1 Introduction

<mark>Introduction</mark> Soil seed banks can be a key part of plant life history strategies that rely on bet hedging and predictive germination to persist in variable environments. Theory and empirical work on seed banks have a long history. Integrating this knowledge into demographic models has proven more challenging. While aboveground individuals can be readily observed, additional experiments are often required to estimate seed survival, viability, and germination. Studies have identified consequences to omitting seed banks from stage-structured models and highlighted the effects of parameter uncertainty associated with dormant stages.

<mark>Conservation implications</mark> Seed banks are generally important for persistence and potentially critical for rare/endangered plant species. Selecting seeds with seed banks can be an important part of restoration strategies <mark>Barga and Leger</mark>. Persistent seed banks may be critical for helping species make it through periods of environmental uncertainty, for desiccation resistant seeds in arid climates experiencing droughts as well as for seeds that must remain moist in tropical climates.

<mark>Seed banks in structured models: matrix structure</mark> Structured demographic models (whether matrix models or integral projection models) require choices about the structure of the life cycle. It's possible to evaluate the consequences of incorrectly specifying the life cycle by, for example, omitting the seed bank stage when it is important. For example, Doak et al. (2002); Nguyen et al. (2019) reviewed published seed bank studies and assessed the effects of removing seed banks on the estimated population growth rate.

<mark>Seed banks in structured models: parameter uncertainty</mark> Parameter uncertainty can also influence estimates of population growth rate. Because relatively little is often known about the seed bank, authors may be less likely to include uncertainty in their assumptions. Paniw

et al. (2017) evaluated the consequences of uncertainty associated with seed bank transitions and quantified the effect of this uncertainty on estimates of population growth rate and population extinction. So, we know that it can be important to include seed banks and that uncertainty about them can be important.

Seed banks in structured models: data-parameter mismatch One challenge is that we are often interested in conditional probabilities; say, the probability of survival or germination for a 2-year old seed. However, because seeds are buried in the soil we can not observe the events leading up to a particular outcome. This idea is closely related to the concept of a hazard in survival analysis Zens and Peart 2003, also referred to as the 'force of mortality'. This reflects the instantaneous death rate. While we may want to observe how this changes over time, we often only have data from the end of a time period.

Focus of this article: new models, model checking An important gap is that there seem to be few ways of checking whether particular representations of seed bank persistence are appropriate. Building a model checking and/or model selection step into the process of constructing stage-structured population models creates the opportunity to consider a richer array of data when modeling seed bank. For example, Peterson et al. (2019) describe model selection for considering beta vs. normal growth.

Motivation: Clarkia We encountered these challenges when we sought to estimate seed survival and germination for population models of the winter annual Clarkia xantiana ssp. xantiana. We thus built on existing approaches by developing and evaluating a hierarchical model that includes a process component to represent loss from the seed bank via decay and germination, and a sampling model to represent the design of the seed burial experiments. Finally, we combined data from estimates of reproduction and seedling emergence in permanent plots with the seed experiments. We used our approach to assess the extent of intraspecific variation in the seed survivorship curve and in age-specific germination for seeds of C. xantiana.

We observed substantial variation in the survivorship curves of seeds in our C. xantiana study populations, corresponding to Type I, II, and III seed survivorship trajectories. The Weibull survival process proved to be a good representation of persistence in the seed bank, and population-level estimates for the shape parameter ranged from .25 to 1.25. We summarized uncertainty about the full survivorship curve as a function of both the shape and inverse-rate parameter. We also propagated uncertainty about survival to estimates of age-specific germination.

In addition to fitting our model to empirical data, we conducted a simulation study to aid with computational implementation. We compared a nonparametric representation of seed survival and several parametric survival functions (negative exponential, continuous exponential, Weibull). We simulated data from each survival process and fit models to confirm we recovered known parameter combinations. We also tested the consequences of fitting a model to data generated under a different process (e.g. fit a model with nonparametric survival to data generated by a Weibull). Overall, we propose that one of the strengths of the approach we describe is that it makes it possible to compare representations of seed decay and germination in the model fitting step before moving on to evaluating its impact on population growth rate.

==Goals== Our approach to analyzing seed bag burial experiments contributes to efforts to make richer inferences from the trove of demographic data collected by plant ecologists. The combination of data from aboveground observations and belowground experiments highlights the value of data fusion techniques even in the case of relatively small datasets. (1) Survey combination of data collection and modeling approaches for seed banks. (2) Describe continuous representations of seed survival in the soil seed bank. (3) Use product integral to combine processes of germination and seed survival. (4) Highlight careful model selection as tool to supplement construction of structured models with seed banks.

==Outline== (Aim 1) Explain product integral and P(S)*P(G) (Aim 2) Describe approach

3

for the simplest case: exponential model, constant germination, ie. unstructured seed bank. Simulate for different values of decay and germination, attempt to recover parameters. Plot 1: show 'true' decay+germination, simulated values, and estimated decay+germination. Plot 2: repeat for a range of k+germination values, from low-high half-life (0-1) to show that sampling variance is highest in the middle, and that uncertainty in parameter estimates can increase at old ages even in unstructured model. Plot 3: true vs. observed parameter values for different sample sizes. (Aim 3) Introduce Weibull model and explain why alternative decay functions might be of interest. Repeat simulations from Aim 2 but only show Plot 2+Plot 3.

2 Literature review

To examine the range of strategies used to model seed banks, we examined data associated with COMPADRE. We identified all matrix models that included seeds or seed banks, and read the associated publications to identify how the authors handled seed stages. Before reading papers for this project, we developed a preliminary coding scheme based on our general understanding of methods in the literature. We considered the following categories for data collection: (1) no data, (2) lab experiments, (3) field experiments, (4) observations, (5) combination. We considered the following categories for representation in models: (1) ignore seed bank, (2) model rates as constant, (3) model rates as random, (4) model rates as function of covariates, (5) model rates with differential equations.

We applied our coding scheme to the literature and then discussed the codings to refine our callings. We summarize our findings in a table (supplement) and in a matrix that cross-tabulates data collection and model representation to understand how ecologists implement methods to study seed banks.

4

- Fox 2001; Failure time analysis chapter, Zens and Peart 2003

- McNair, Seed Science Research 2012. Discuss non-parametric and semi-parametric methods for seed germination data.

- Paniw et al. 2016 describe the consequences of parameter uncertainty in dormant life stages on stochastic demographic models.

- Meyer et al. 2006 12 year seed bank experiment with Lepidium

- Menges 2000 reviews plant PVAs

- Satterthwaite et al. 2007 conservation application

- Peterson et al. 2019 discusses improving structured models with representations of non-normal growth; similar logic could be applied to seed bank.

- Bartholomew et al. 1973 and Lewis 1962 discuss catastrophic selection and lack of seed bank in Clarkia.

- Ecology of Soil Seed Banks: book describing the soil seed bank, includes chapter by Venable on evolutionary ecology.

- Doak et al. 2002 describes challenges of modeling seed banks

3 Models

Seed banks present a challenge for models of plant population demography. Individual seeds can not be tracked in the field so ecologists have turned to various methods to assess survival and germination from the seed bank, including experimental burials and seed addition experiments. This makes seed banks a good case for combining observation and process models in order to separate variation due to sampling and variation due to the process in question.

While inputs to the soil seed bank come solely from seed rain, seeds can leave the soil seed bank through various routes. These may include predation, pathogens, deep burial, physiological death, failed germination, and successful germination Simpson et al. Ch. 1 Ecology of seed banks; Figure 1. Seeds can be destroyed by predators or decomposers, lose

viability through natural causes (i.e. age-related loss of vitality) ==Baker Chapter 2 Ecology of seed banks==

See ==Baskin and Baskin Chapter X Ecology of seed banks== for description of progression of dormancy. See ==McCue and Holtsford 1998== for description of how dormancy progresses over time in Clarkia.

Seeds leave the soil seed bank through destruction, physiological death, and germination. The persistence of seeds in the soil combines discrete and continuous processes. Seeds leave the seed bank through germination (discrete) or destruction (continuous). The persistence of seeds to the first winter combines destruction and physiological death. However, in January we only directly observe seed loss due to destruction. We also observe loss through germination. Seeds are then buried again and unearthed in October at which point we observe additional loss due to destruction. We also estimate physiological death at that time. We estimate loss due to destruction at January. We estimate loss due to germination at January. We then estimate loss due to destruction in October, conditional on not being destroyed or germinating in January. Finally, we then estimate loss due to physiological death in October, conditional on not being destroyed. We want to estimate the loss due to physiological death, conditional on not germinating and not being destroyed in January;

Loss of seeds from the soil seed bank reflect the product of destruction and germination. Seeds that persist to a given time are those that have not been destroyed and did not germinate. Additionally, we estimate the proportion of seeds that persist to October of each year that are viable. However, these tests are conducted on a subset of seeds; we can say viability in October — intact in October. We are interested in the proportion of seeds that did not germinate and are intact in January that are viable. To get this number requires some additional assumptions. The product of persistence and viability gives the expected proportion of seeds viable in October.

Here is another option. I could estimate viability in October and then calculate the

proportion of seeds that are viable in January using either an exponential decay OR assuming decay proceeds at the same rate as physical death. In either case, we just use data from each age to estimate the viability of seeds of that age, rather than inferring the viability of seeds at all points in time.

Basically, describe two different events (physical destruction) and (germination). The curve for physical destruction is the product of seed decay and germination. Germination is discrete and includes survival to time x, viability at time x, and then germination.

The persistence of seeds beyond the first germination time is thus conditional on the absence of germination. Both germination and mortality remove seeds from the soil seed bank; we define hazard functions for these events. The survival function thus describes the persistence of seeds in the soil (seeds that do not experience mortality or germination).

The survival function is defined by the product integral of the discrete and continuous hazards:

$$
\begin{aligned}
\Pr(B|A) &= \frac{\Pr(A|B)\Pr(B)}{\Pr(A)} \\
\Pr(g|s) &= \frac{\Pr(s|g)\Pr(g)}{\Pr(s)} \\
\Pr(g) &= Pr(g|s)\Pr(s)
\end{aligned}
\tag{1}
$$

Finally, seeds can also leave the soil seed bank through loss of viability. We expect that loss of viability is a continuous process but our estimates of viability are discrete.

For the deterministic process model, I turned to literature on litter decomposition e.g. Olson 1963, Cornwell et al. 2014. Cornwell et al. 2014 describe several functions that I considered as "process models". Some of these are the same as the survival functions underlying the models in Rees and Long 1993. There are several models that could be considered to represent the survival process that I briefly describe below. A separate study with different objectives might be interested in coding up these competing models and using model

7

selection to evaluate the relative support for different models of survival.

The negative exponential model corresponds to a constant decay rate, as in Gremer and Venable 2014. This model has a constant decay parameter k that implies the decaying material be treated as a homogeneous mass with each seed having an equal probability of mortality throughout time. The negative exponential decay model may be appropriate if seed survival is random in time and largely unrelated to seed characteristics. But we know that seeds exhibit phenotypic variation in dormancy-related traits Lonsdale 1988.

The continuous exponential model is one in which seed mortality is described by a continuous quality distribution. For example, if seed survival is a function of a continuous phenotypic trait, we could imagine this model describing the survival process.

The Weibull residence time model represents the seed pool as a distribution of survival times. Parameter alpha controls the shape of the decomposition trajectory and beta the rate of decomposition. This model can reduce to the exponential model when alpha is 1. If alpha is less than 1, the decomposition rate decreases through time. If alpha is greater than 1, the decomposition rate increases through time.

The table below summarizes the models we considered, the parameters, and the number of parameters for each model. The product of the mortality and germination process is a deterministic function that equals the average proportion of seeds that are expected to be intact or germinated. This average proportion is the mean of a beta distribution with process variance that captures variation beyond the effect of seed age on mortality as represented with a mortality process and germination. We might also consider a germination process where germination is randomly distributed and independent for each bag?

Table 1: Deterministic process models for seed mortality and germination

8

Model	Mortality process	Germination process	Parameters	Parameter number
A1	Negative exponential	Constant	k, g	2
A2	Negative exponential	Age-dependent	k, g_{1-3}	4
B1	Compound exponential	Constant	a, b, g	3
B2	Compound exponential	Age-dependent	a, b, g_{1-3}	5
C1	Weibull residence time	Constant	α, β, g	3
C2	Weibull residence time	Age-dependent	α, β, g_{1-3}	5

3.1 Linking to seed pots

The seed bag burial experiment made it possible to estimate conditional germination; that is, the probability of germination for seeds of a given age that were intact at a given time. In the seed pot experiment, observations of seedlings reflect unconditional germination; the probability of emergence for a seedling that was buried some time ago. This process reflects both seed survival and conditional germination, but we only observe the emergence outcome.

In the model defined above, survival has been reduced to a two-parameter process: roughly, a scale and shape parameter. I estimated a population-level shape parameter (α_j) that defines how the probability of survival changes through time. For each population, I estimated a scale parameter (η_{jk}) for each year; this describes the magnitude of change and essentially shifts the survival curve up or down. The posteriors from these estimates could serve as informed priors.

For the seed pot data, we only have observations on unconditional germination. The latent probability would thus be the product of survival up to the time of germination and conditional germination. This amounts to changing the equations for the compound event histories to reflect the events under observation. The germination probabilities are indexed by site j, experimental year k, and age l; for example, emergence of 1-year old seeds at site j from the experimental year k is given by γ_{jkl}. The compound event histories are now defined as:

9

$$\theta_{j1,1} = \gamma_{j1,1}$$
$$\theta_{j1,2} = (1 - \gamma_{j1,1}) \times \gamma_{j1,2} \qquad (2)$$
$$\theta_{j1,3} = (1 - \gamma_{j1,1}) \times (1 - \gamma_{j1,2})\gamma_{j1,3}$$

In each experimental year k, we have data on emergence 1, 2, and 3 years after the seeds were buried. Multiplying these with the survival functions returns the unconditional probability of germination, which is the latent probability underlying the binomial likelihood for the observed counts of seedlings.

We would use the estimates for the parameters of the survival functions from the seed bags as informed priors. The posteriors on parameters of the survival function would remain the same if the data from the seed pot experiments was consistent with the estimates from after the seed bag experiments. They would change if additional information from the seed pot experiments influenced the estimate. Similarly, the population-level conditional germination probabilities could also serve as informed priors.

3.2   Model selection

For the deterministic process model, I turned to literature on litter decomposition e.g. Olson 1963, Cornwell et al. 2014. Cornwell et al. 2014 describe several functions that I considered as "process models". Some of these are the same as the survival functions underlying the models in Rees and Long 1993. There are several models that could be considered to represent the survival process that I briefly describe below. A separate study with different objectives might be interested in coding up these competing models and using model selection to evaluate the relative support for different models of survival.

4 Overview

₂₂₅ Seed banks present a challenge for models of plant population demography. Individual seeds

₂₂₆ can not be tracked so ecologists have turned to various methods to assess survival and ger-

₂₂₇ mination from the seed bank, including experimental burials and seed addition experiments.

₂₂₈ The base assumption is often that seed mortality proceeds at a constant rate. Although this

₂₂₉ assumption has repeatedly been challenged Lonsdale 1988, Rees and Long 1993 there remain

₂₃₀ limited assessments for how appropriate it is.

₂₃₁ One challenge is that we are often interested in conditional probabilities; say, the prob-

₂₃₂ ability of survival or germination for a 2-year old seed. However, because seeds are buried

₂₃₃ in the soil we can not observe the events leading up to a particular outcome. This idea

₂₃₄ is closely related to the concept of a hazard in survival analysis Zens and Peart 2003, also

₂₃₅ referred to as the 'force of mortality'. This reflects the instantaneous death rate. While we

₂₃₆ may want to observe how this changes over time, we often only have data from the end of a

₂₃₇ time period.

Conditional versus unconditional probability Christian and Stanton 2004, Vitalis et al. 2004, Rees and Long 1993, Rees and Long 1992, Kalisz 1991, Templeton and Levin 1979, Venable 1989. Conditional probabilities of survival and germination are conditional on survival up to a point.

$$
\begin{aligned}
\Pr(B|A) &= \frac{\Pr(A|B)\Pr(B)}{\Pr(A)} \\
\Pr(g|s) &= \frac{\Pr(s|g)\Pr(g)}{\Pr(s)} \\
\Pr(g) &= Pr(g|s)\Pr(s)
\end{aligned}
\tag{3}
$$

₂₃₈ Three works illustrate some of the range of approaches that have been used to study seed

₂₃₉ dynamics. Kalisz 1991 describes a seed burial experiment in which known numbers of seeds

were buried and unearthed after varying periods of time. Kalisz estimates unconditional and conditional survival and emergence probabilities over the course of two years. Rees and Long 1993 present an analysis of experimental seed burials that focuses on seedling recruitment. They fit several distributions to 5 years of data on emerged seedlings, and use goodness-of-fit tests to assess the survival functions they consider. Gremer and Venable 2014 estimate seed survival by assuming a constant mortality rate throughout the year for new seeds and one-year old seeds (Gremer and Venable Appendix S1). Their study uses estimates of seed production and soil cores to infer the rate of survival.

Our seed bag experiment combines some features of these studies. We used seed bags buried with known numbers of seeds to estimate seedling emergence and survival as a function of age in three consecutive years. We thus have data on unconditional survival and emergence probabilities. Sampling seed bags at the end of each experimental year was also destructive. And as is generally the case we do not have independent information about survivorship patterns in the seed bank that could inform whether or not seed mortality is constant.

Here, I give a general overview of the approach we used to understand the dynamics of seeds. Although the full model is hierarchical, I ignore the hierarchical component in my explanation at this point. Because we have data on both seed survival and germination, we developed models that used both parts of the experiment. For seed bags that were unearthed in February, the field crew counted the germinants as well as the intact, ungerminated seeds. I used these counts to estimate the conditional probability of germination for 1-, 2-, or 3-year old seeds in each experimental round.

The sum of counts of germinants and intact, ungerminated seeds in February of a given year correspond to the unconditional probability of surviving to that time. Likewise, the counts of intact seeds in October of a given year correspond to the unconditional probability of surviving to that time. If individual bags were followed over the course of the entire experiment, it might be possible to estimate the probability of survival between time points

because it would be unlikely (or impossible) that bags would gain seeds over the course of the experiment. However, sampling in October of each experimental year was destructive and so seed counts increased between time points because of sampling variation.

To represent the experimental design and model the survival process, I developed a model with a binomial likelihood (the sampling model) and a deterministic process model that represents survival. Because we don't know which model best represents the survival process in our seeds, I chose to work with the Weibull residence time model. The Weibull is a flexible model that can represent a range of decay/survival trajectories Cornwell et al. 2014, is used in survival analysis Klein and Moeschberger 2003; Pinder et al. 1978, and has been applied to other problems in ecology and evolutionary biology Smits 2015; Dahlgren et al. 2016. Fitting a model with a Weibull survival process doesn't preclude the possibility that seed mortality is constant; the negative exponential model is a subset.

Crucially, seeds in the seed bags are subject to these two separate processes: germination and survival. For example, to estimate the probability that a seed survives to October of year two, we need to account for both the probability that the seed remains intact (does not suffer mortality) and the probability that the seed does not germinate. I linked the germinant and seed counts to jointly estimate the survival and germination probability; the estimates of survival thus incorporate information about the probability of germination.

Finally, seed production happens in June/July but the seed bags were buried in October. To estimate survival between seed production and the start of the seed bag experiment, I used estimates of seed production and seedling emergence in the permanent plots to Eckhart et al. 2011. In two experimental years, we had data on the number of seeds entering a plot and the number of seeds emerging alongside the seed bag experiment. I summed counts across all plots and wrote an additional likelihood statement for the probability of observing those counts. The additional likelihood composed the unobserved probability of surviving from seed production to October with the probability of survival from October to February

(given by the deterministic survival process) and the probability of germination.

For the deterministic process model, I turned to literature on litter decomposition e.g. Olson 1963, Cornwell et al. 2014. Cornwell et al. 2014 describe several functions that I considered as "process models". Some of these are the same as the survival functions underlying the models in Rees and Long 1993. There are several models that could be considered to represent the survival process that I briefly describe below. A separate study with different objectives might be interested in coding up these competing models and using model selection to evaluate the relative support for different models of survival.

The negative exponential model corresponds to a constant decay rate, as in Gremer and Venable 2014. This model has a constant decay parameter k that implies the decaying material be treated as a homogeneous mass with each seed having an equal probability of mortality throughout time. The negative exponential decay model may be appropriate if seed survival is random in time and largely unrelated to seed characteristics. But we know that seeds exhibit phenotypic variation in dormancy-related traits (?).

The continuous exponential model is one in which seed mortality is described by a continuous quality distribution. For example, if seed survival is a function of a continuous phenotypic trait, we could imagine this model describing the survival process.

The Weibull residence time model represents the seed pool as a distribution of survival times. Parameter alpha controls the shape of the decomposition trajectory and beta the rate of decomposition. This model can reduce to the exponential model when alpha is 1. If alpha is less than 1, the decomposition rate decreases through time. If alpha is greater than 1, the decomposition rate increases through time.

The table below summarizes the models we considered, the parameters, and the number of parameters for each model. The product of the mortality and germination process is a deterministic function that equals the average proportion of seeds that are expected to be intact or germinated. This average proportion is the mean of a beta distribution with process

14

318 variance that captures variation beyond the effect of seed age on mortality as represented

319 with a mortality process and germination. We might also consider a germination process

320 where germination is randomly distributed and independent for each bag?

Table 2: Seed bag dataset models

Model	Mortality process	Germination process	Parameters	Parameter number
A1	Negative exponential	Constant	k, g	2
A2	Negative exponential	Age-dependent	k, g_{1-3}	4
B1	Compound exponential	Constant	a, b, g	3
B2	Compound exponential	Age-dependent	a, b, g_{1-3}	5
C1	Weibull residence time	Constant	α, β, g	3
C2	Weibull residence time	Age-dependent	α, β, g_{1-3}	5

322 The table above shows the survival, probability mass, and hazard functions for the persis-

323 tence of seeds in the soil. It also shows the survival, probability mass, and hazard functions

324 for viable seeds in the soil. Because the event X here is the mortality of seeds, the proba-

325 bility of survival over an interval is the complementary probability to the hazard function.

326 Seeds that physically persist in the soil may be physiologically dead, so this is closer to

327 apparent survival as in the mark-recapture literature. The table above can be translated

328 into parameters used in the stage-structured matrix model as follows:

Table 3: Structured model parameters conditional on persistence only

Matrix parameter	Probability statement	Probability
g_1	$P(G_1\|S_1)$	γ_1
g_2	$P(G_2\|S_3, S_2, G_1^c, S_1)$	γ_2
g_3	$P(G_3\|S_5, S_4, G_2^c S_3, S_2, G_1^c, S_1)$	γ_3
s_1	$P(S_1)$	θ_1
s_2	$P(S_2\|G_1^c, S_1)$	θ_3/θ_2
s_3	$P(S_3\|S_2, G_1^c, S_1)$	θ_4/θ_3
s_4	$P(S_4\|G_2^c S_3, S_2, G_1^c, S_1)$	θ_6/θ_5
s_5	$P(S_5\|S_4, G_2^c S_3, S_2, G_1^c, S_1)$	θ_7/θ_6
s_6	$P(S_6\|G_3^c, S_5, S_4, G_2^c S_3, S_2, G_1^c, S_1)$	θ_9/θ_8

While inputs to the soil seed bank come solely from seed rain, seeds can leave the soil seed bank through various routes. These may include predation, pathogens, deep burial, physiological death, failed germination, and successful germination Simpson et al. Ch. 1 Ecology of seed banks; Figure 1. Seeds can be destroyed by predators or decomposers, lose viability through natural causes (i.e. age-related loss of vitality) Baker Chapter 2 Ecology of seed banks

See Baskin and Baskin Chapter X Ecology of seed banks for description of progression of dormancy. See McCue and Holtsford 1998 for description of how dormancy progresses over time in Clarkia.

Seeds leave the soil seed bank through destruction, physiological death, and germination. The persistence of seeds in the soil combines discrete and continuous processes. Seeds leave the seed bank through germination (discrete) or destruction (continuous). The persistence of seeds to the first winter combines destruction and physiological death. However, in January we only directly observe seed loss due to destruction. We also observe loss through germination. Seeds are then buried again and unearthed in October at which point we observe additional loss due to destruction. We also estimate physiological death at that time. We estimate loss due to destruction at January. We estimate loss due to germination at January. We then estimate loss due to destruction in October, conditional on not being destroyed or germinating in January. Finally, we then estimate loss due to physiological death in October, conditional on not being destroyed. We want to estimate the loss due to physiological death, conditional on not germinating and not being destroyed in January;

Loss of seeds from the soil seed bank reflect the product of destruction and germination. Seeds that persist to a given time are those that have not been destroyed and did not germinate. Additionally, we estimate the proportion of seeds that persist to October of each year that are viable. However, these tests are conducted on a subset of seeds; we can say viability in October — intact in October. We are interested in the proportion of seeds that did not germinate and are intact in January that are viable. To get this number requires some additional assumptions. The product of persistence and viability gives the expected proportion of seeds viable in October.

Here is another option. I could estimate viability in October and then calculate the proportion of seeds that are viable in January using either an exponential decay OR assuming decay proceeds at the same rate as physical death. In either case, we just use data from each age to estimate the viability of seeds of that age, rather than inferring the viability of seeds at all points in time.

Basically, describe two different events (physical destruction) and (germination). The curve for physical destruction is the product of seed decay and germination. Germination is discrete and includes survival to time x, viability at time x, and then germination.

The persistence of seeds beyond the first germination time is thus conditional on the absence of germination. Both germination and mortality remove seeds from the soil seed bank; we define hazard functions for these events. The survival function thus describes the persistence of seeds in the soil (seeds that do not experience mortality or germination).

The survival function is defined by the product integral of the discrete and continuous

hazards:

$$\Pr(B|A) = \frac{\Pr(A|B)\Pr(B)}{\Pr(A)}$$
$$\Pr(g|s) = \frac{\Pr(s|g)\Pr(g)}{\Pr(s)} \tag{4}$$
$$\Pr(g) = Pr(g|s)\Pr(s)$$

Finally, seeds can also leave the soil seed bank through loss of viability. We expect that loss of viability is a continuous process but our estimates of viability are discrete.

Overview

We first describe the experimental design used to estimate belowground vital rates and illustrate how we combine estimates from a seed bag burial experiment and viability trials to obtain parameters for a structured population model (Section 5). We use a conceptual diagram to connect the experimental design and data (Figure 6), and describe how estimates from the data are used to compose parameters for a structured population model (Tables 4 and 5).

We then describe the data and model construction for the seed bag experiment (Section 6). This section also includes the directed acyclic graphs (DAGs) for any likelihoods associated with the model construction. The section also describes how we estimated survival from reproduction to October when the seed bag experiments started.

We then describe the data and model construction for the viability trials and again include the DAGs (Section 7).

Finally, Section 8 describes how we use the parameter estimates from Section 6 and 7 to obtain the parameters for the structured population model from Section 5.

==Note to self about how to turn this into Methods.== The key pieces are Figure 6, columns 1, 3, and 6 of Table 4 and columns 1 & 3 in Table 5. Figure 6 illustrates the relationship between the seed bag burial experiment, viability trials, and survival/germination probabilities. Table 4 gives the mathematical relationship among the components of Figure 6. Table 5 translates the survival and germination components to parameters for the structured population model. I would to pare this down I would pare down Section 5 to describe the experiment and reduce the tables to the relevant parameters, likely combining them into 1 table (top section being the survival functions, bottom section being the model parameters). I would keep the description of the full model and the viability models. I would move any discussion of how the models were constructed and DAGs to a separate document. I would reduce the section on translating estimated parameters to model parameters to generally describe how this was done and move the rest to a separate document.

5 Experimental design and parameters

₃₉₉ We used a field experiment in which we buried known numbers of seeds to estimate the
₄₀₀ probability of germination and survival for seeds of different ages (Figure 6A, gray panel).
₄₀₁ Buried bags were unearthed two times each during their first, second, or third year. Bags
₄₀₂ that were dug up in the first year were only used to count intact seeds during the first year;
₄₀₃ at the end fo the year bags were removed fro the field. The seed bag burial experiment was
₄₀₄ initiated in 3 consecutive years (3 rounds) at the 20 study populations. In round 1, bags
₄₀₅ were dug up in year 1, 2, and 3. In round 2, bags were dug up in year 1 and 2. In round 3,
₄₀₆ bags were dug up in year 1. There were thus 3 sets of counts for age 1 seeds, 2 sets of counts
₄₀₇ for age 2 seeds, and 1 set of counts for age 3 seeds.

₄₀₈ In summary, during each experimental round we collected data on the number of seeds in
₄₀₉ the soil seed bank for up to 3 years. These estimates correspond to the number of seeds that
₄₁₀ remain in the soil seed bank. We thus define a persistent seed as one that remains intact in the
₄₁₁ seed bags (i.e. was countable in the seed bags); seeds were lost through physical destruction,
₄₁₂ decay, or germination. In (Figure 6A), points represent the average seed numbers associated
₄₁₃ with hypothetical counts in seed bags across an experimental round.

₄₁₄ At the end of each experimental year, bags were brought to the lab and intact seeds
₄₁₅ were tested in a two-stage viability trial (Figure 6B, gray panel). In the lab, we conducted
₄₁₆ germination trials and viability assays on subsets of the seeds from each bag to estimate the
₄₁₇ viability of the intact seeds. First, we placed up to 15 seeds from each bag on to moist filter
₄₁₈ paper in a disposable cup and observed germination over 10 days; we counted and removed
₄₁₉ germinants every 2 days. After 10 days, all remaining ungerminated seeds (up to a total
₄₂₀ of 10 seeds) were sliced in half and individually placed into the wells of 96-well plates filled
₄₂₁ with a solution of tetrazolium chloride, which stains viable tissue red. [Eckhart et al. 2011:
₄₂₂ not all ungerminated seeds were tested; most were] We covered the plates with foil. Each
₄₂₃ 96-well plate contained seed from at least one bag per population of a given seed-age class.
₄₂₄ Two or three tests of up to 15 seeds each were conducted for each bag. We checked and
₄₂₅ counted for viable seeds every 2 days for 10 days. We conducted viability trials in each year
₄₂₆ we conducted seed burial experiments.

₄₂₇ We tested the viability of seeds in October, and were thus able to estimate the proportion
₄₂₈ of viable seeds (Figure 6B; filled points). We further inferred the viability of intact seeds
₄₂₉ in January by assuming that seeds lost viability at a constant rate (exponential decay).
₄₃₀ Further, we interpolated between estimates by assuming that viability changed at a constant
₄₃₁ rate between years, and that all seeds were viable at the start of the experiment (Figure 6B;
₄₃₂ open points).

₄₃₃ We used the data from the seed bag burial experiment and viability trials to estimate age-
₄₃₄ specific seed survival and germination. We start with survival. To construct our estimates
₄₃₅ of seed survival, we first describe the discrete survival functions, probability mass functions,
₄₃₆ and hazards (e.g. Klein and Moeschberger 2003) associated with the persistence and viability
₄₃₇ of seeds in the soil seed bank (Table 4). We define a persistent seed as one that remains
₄₃₈ intact in the seed bags and a viable seed as one that is intact and viable (Thompson et al.

439 2003?).

Table 4: Seed persistence and viability in the soil seed bank

Time	x_i	Persistence			Viability		
		$S(x_i)$	$p(x_i)$	$h(x_j)$	$S(x_i)$	$p(x_i)$	$h(x_j)$
Oct_0	x_0	θ_0	1	—	$\phi_0 = \theta_0$	1	—
$\mathrm{Jan}_{1,\mathrm{total}}$	x_1	θ_1	$1 - \theta_1$	$\frac{1-\theta_1}{1}$	$\phi_1 = \theta_1(\gamma_1 + (1-\gamma_1)\nu_1^{1/3})$	$1 - \phi_1$	$\frac{1-\phi_1}{1}$
$\mathrm{Jan}_{1,\mathrm{intact}}$	x_2	θ_2	$\theta_1 - \theta_2$	$\frac{\theta_1-\theta_2}{\theta_1}$	$\phi_2 = \theta_2\nu_1^{1/3}$	$\phi_1 - \phi_2$	$\frac{\phi_1-\phi_2}{\phi_1}$
Oct_1	x_3	θ_3	$\theta_2 - \theta_3$	$\frac{\theta_2-\theta_3}{\theta_2}$	$\phi_3 = \theta_3\nu_1$	$\phi_2 - \phi_3$	$\frac{\phi_2-\phi_3}{\phi_2}$
$\mathrm{Jan}_{2,\mathrm{total}}$	x_4	θ_4	$\theta_3 - \theta_4$	$\frac{\theta_3-\theta_4}{\theta_3}$	$\phi_4 = \theta_4(\gamma_2 + (1-\gamma_2)\nu_1(\nu_2/\nu_1)^{1/3})$	$\phi_3 - \phi_4$	$\frac{\phi_3-\phi_4}{\phi_3}$
$\mathrm{Jan}_{2,\mathrm{intact}}$	x_5	θ_5	$\theta_4 - \theta_5$	$\frac{\theta_4-\theta_5}{\theta_4}$	$\phi_5 = \theta_5\nu_1(\nu_2/\nu_1)^{1/3}$	$\phi_4 - \phi_5$	$\frac{\phi_4-\phi_5}{\phi_4}$
Oct_2	x_6	θ_6	$\theta_5 - \theta_6$	$\frac{\theta_5-\theta_6}{\theta_5}$	$\phi_6 = \theta_6\nu_2$	$\phi_5 - \phi_6$	$\frac{\phi_5-\phi_6}{\phi_5}$
$\mathrm{Jan}_{3,\mathrm{total}}$	x_7	θ_7	$\theta_6 - \theta_7$	$\frac{\theta_6-\theta_7}{\theta_6}$	$\phi_7 = \theta_7(\gamma_3 + (1-\gamma_3)\nu_2(\nu_3/\nu_2)^{1/3})$	$\phi_6 - \phi_7$	$\frac{\phi_6-\phi_7}{\phi_6}$
$\mathrm{Jan}_{3,\mathrm{intact}}$	x_8	θ_8	$\theta_7 - \theta_8$	$\frac{\theta_7-\theta_8}{\theta_7}$	$\phi_8 = \theta_8\nu_2(\nu_3/\nu_2)^{1/3}$	$\phi_7 - \phi_8$	$\frac{\phi_7-\phi_8}{\phi_7}$
Oct_3	x_9	θ_9	$\theta_8 - \theta_9$	$\frac{\theta_8-\theta_9}{\theta_8}$	$\phi_9 = \theta_9\nu_3$	$\phi_8 - \phi_9$	$\frac{\phi_8-\phi_9}{\phi_8}$

441 The discrete survival function for persistent seeds (Table 4) represents seeds that do not

442 leave the soil seed bank, either through physical destruction or germination, and summarizes

443 the survival at marked intervals in the graph in (Figure 6A). The discrete survival function for

444 viable seeds (Table 4) represents seeds that do not leave the soil seed bank through physical

445 destruction, germination, or physiological death. To obtain the second survival function, we

446 multiplied the probability of persistence at each time by the estimated or inferred viability.

447 We also calculated the proportion of viable seeds before germination in January by assuming

448 that all seedlings were viable before germination. The survival function for persistent seeds

449 sets an upper bound for the survival function of viable seeds. Lower estimates of viability

450 decrease the survival function of viable seeds, with effects that compound over time (Figure

451 6D).

452 The survival function for viable seeds (ϕ) is composed of estimates of persistence over

453 time ($\theta_.$), estimates of viability ($\nu_.$), and estimates of germination conditional on persistence

454 ($\gamma_.$). We used the survival function and germination probabilities to define the parameters

455 in the structured population model for *Clarkia xantiana* ssp. *xantiana*. Table 5 defines

19

the age-specific germination probabilities and survival probabilities for the Leslie matrix in <mark>Eckhart et al. 2011</mark> in terms of the survival function (Table 4) and germination probabilities.

Table 5: Structured model parameters conditional on persistence and viability

Parameter	Probability statement	Probability
g_1	$P(G_1 \mid S_1)$	γ_1/ϕ_1
g_2	$P(G_2 \mid S_3, S_2, G_1^c, S_1)$	γ_2/ϕ_4
g_3	$P(G_3 \mid S_5, S_4, G_2^c S_3, S_2, G_1^c, S_1)$	γ_3/ϕ_7
s_1	$P(S_1)$	ϕ_1
s_2	$P(S_2 \mid G_1^c, S_1)$	ϕ_3/ϕ_2
s_3	$P(S_3 \mid S_2, G_1^c, S_1)$	ϕ_4/ϕ_3
s_4	$P(S_4 \mid G_2^c S_3, S_2, G_1^c, S_1)$	ϕ_6/ϕ_5
s_5	$P(S_5 \mid S_4, G_2^c S_3, S_2, G_1^c, S_1)$	ϕ_7/ϕ_6
s_6	$P(S_6 \mid G_3^c, S_5, S_4, G_2^c S_3, S_2, G_1^c, S_1)$	ϕ_9/ϕ_8

Figure 6C illustrates the relationship among the pieces discussed here. Estimates of germination from the seed bag experiment correspond to the probability of germination conditional on persistence (e.g. γ_1). Multiplying these estimates by the probability of persistence in January before germination gives the unconditional probability of germination (e.g. $\theta_1 \times \gamma_1$). Finally, the probability of germination conditional on viability is estimated by incorporating loss of viability into the survival function (e.g. γ_1/ϕ_1).

6 Seed bag burial experiments

6.1 Data

We collected the following data from the seed bag burial experiments:

- $n_{g,ijkl}$ = observed count of intact seeds and seedlings in the seed bags in January in the ith bag, from the jth population, from the kth experimental year, assumed to be measured perfectly

- $y_{g,ijkl}$ = observed count of germinated seedlings in the seed bags in January in the ith bag, from the jth population, from the kth experimental year, for seeds of age l, assumed to be measured perfectly

- n_{ijk} = observed count of seeds in the seed bags at the start of the experiment in October in the ith bag, from the jth population, from the kth experimental year, assumed to be measured perfectly

- y_{ijk} = observed count of ungerminated, total seeds in the seed bags in January (ungerminated seeds plus seedlings) and October in the ith bag, from the jth population, from the kth experimental year, assumed to be measured perfectly

We started these seed burial experiments in three subsequent years (2005, 2006, 2007) to obtain multiple counts to estimate for seed survival and germination. These are two sets of data; the data with subscript g are used to estimate conditional germination and the data without a subscript are used to estimate survival.

6.2 Models

I build the models for each dataset before linking them. First, I describe a model for the age-specific, conditional germination probability. Second, I describe a model for the probability of seed survival.

6.2.1 Conditional germination

The joint likelihood for the germination model is given by

$$
\begin{aligned}
[\boldsymbol{\mu}_\mathrm{g}, \boldsymbol{\sigma}_\mathrm{g}, \boldsymbol{\mu}_\mathrm{g}^\mathrm{pop}, \boldsymbol{\sigma}_\mathrm{g}^\mathrm{pop} | \mathbf{y}_\mathrm{g}] \propto & \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \prod_{l=1}^{L} \mathrm{binomial}(y_{\mathrm{g},ijkl} | n_{\mathrm{g},ijkl}, \mathrm{logit}^{-1}(\alpha_{\mathrm{g},ijkl})) \\
& \times \mathrm{normal}(\alpha_{\mathrm{g},ijkl} | \mu_{\mathrm{g},jkl}, \sigma_{\mathrm{g},jkl}) \\
& \times \mathrm{normal}(\mu_{\mathrm{g},jkl} | \mu_{\mathrm{g},jl}^\mathrm{pop}, \sigma_{\mathrm{g},jl}^\mathrm{pop}) \\
& \times \mathrm{half\text{-}normal}(\sigma_{\mathrm{g},jkl} | 0, 1) \\
& \times \mathrm{normal}(\mu_{\mathrm{g},jl}^\mathrm{pop} | 0, 1) \mathrm{half\text{-}normal}(\sigma_{\mathrm{g},jl}^\mathrm{pop} | 0, 1).
\end{aligned}
\tag{5}
$$

in which J indexes the 20 populations in the study; K indexes the 3 experimental years; L indexes the 3 ages; I indexes the observations within each set of data. The model has a binomial likelihood and logit-link for the latent probability of germination. The latent probability of germination at each age is described by two hierarchical levels: the first level is the experimental years and the second (upper) level is the population. The variance parameters have weakly informative, half-normal priors and the mean parameters have weakly informative, normal priors. See Appendix: Priors for details on the choice of weakly informative priors.

6.2.2 Survival

I represented the seed mortality process using a deterministic function, the Weibull survival function. Smits 2015 uses the following parameterization of the Weibull survival function:

$$
S(t) = \exp\left(-\left(\frac{t}{\beta}\right)^\alpha\right)
\tag{6}
$$

for which β is corresponds to the inverse of the scale parameter in the exponential distribution, accounting for the influence of the shape parameter. Formally, if λ is the scale parameter, $\lambda = \beta^{-\frac{1}{\alpha}}$ so that $\beta = 1/\lambda^\alpha$ ==Smits 2015, ProbOnto==. To model the influence of group effects (random intercepts $\eta_i j$), $\beta_i j = \exp(-\frac{\eta_i j}{\alpha})$. This approach would also extend to incorporating covariates into the numerator to model their effect on the scale parameter.

Parameter alpha controls the shape of the decomposition trajectory and beta the rate of decomposition. This model can reduce to the exponential model when alpha is 1. If alpha is less than 1, the decomposition rate decreases through time. If alpha is greater than 1, the decomposition rate increases through time.

Counts of seeds were indexed by sites j, rounds of observation starting in a series of years k, and observation m. Here, the observation index tracks the time at which a particular observation. The joint likelihood for the survival model is then given by

$$\beta_{ijk} = \exp(-\frac{\eta_{ijk}}{\alpha_j})$$

$$g(\eta_{ijk}, \boldsymbol{\theta}, t_{ijkm}) = \boldsymbol{\theta} \exp(-(\frac{t_{ijkm}}{\beta_{ijk}})^{\alpha_j})$$

$$= \boldsymbol{\theta} \exp(-(\frac{t_{ijkm}}{\exp(-\frac{\eta_{ijk}}{\alpha_j})})^{\alpha_j})$$

$$[\boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\mu}^{\mathrm{pop}}, \boldsymbol{\sigma}^{\mathrm{pop}}|\mathbf{y}] \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\prod_{m=1}^{M} \mathrm{binomial}(y_{ijkm}|n_{ijkm}, g(\boldsymbol{\theta}, \eta_{ijk}, \alpha_j, t_{ijkm}))$$

$$\times \mathrm{normal}(\eta_{ijk}|\mu_{jk}, \sigma_{jk})$$

$$\times \mathrm{normal}(\mu_{jk}|\mu_j^{\mathrm{pop}}, \sigma_j^{\mathrm{pop}})$$

$$\times \text{half-normal}(\sigma_{jk}|0, 1)$$

$$\times \mathrm{normal}(\mu_j^{\mathrm{pop}}|0, 1)\text{half-normal}(\sigma_j^{\mathrm{pop}}|0, 1)$$

$$\times \mathrm{gamma}(\alpha_j|2, 2).$$

(7)

The Weibull has a shape and scale parameter; we estimate a shape parameter for each

population (α_j) <mark>Smits 2015</mark>. This is equivalent to assuming the rate of change in survivorship is a population-level property but that the scale varies from year to year within each population (η_{ijk}).

The survival data has to account for the loss of seeds due to mortality as well as germination. Because we model germination as conditionally dependent on the number of intact seeds at the time of germination, we estimate the germination rate independently. We combine that information with the survival time model. We use the survival function associated with a Weibull distribution as the deterministic function for how seed survivorship changes over time. In the likelihood above, I write the influence of germination as $\boldsymbol{\theta}$ and describe how this connects to the germination model next.

6.2.3 Germination histories

I represent the germination history of seeds in the experiment with distributions γ_{ya}, describing the probability that a seed has germinated. The subscript corresponds to seeds of age $a = \ldots$ and experimental year $y = \ldots$.

The experiment was initiated in three consecutive years (2006, 2007, 2008). Seed bags from the first year were collected at ages 1, 2, and 3. Seed bags from the second year were collected at ages 1 and 2. Seed bags from the third year were collected at age 3. There are thus the following probability distributions:

$$\gamma_{1a} = (\gamma_{11}, \gamma_{12}, \gamma_{13})$$
$$\gamma_{2a} = (\gamma_{21}, \gamma_{22}) \tag{8}$$
$$\gamma_{3a} = (\gamma_{31})$$

I used these probability distributions to summarize the event histories leading up to each count. In total, I defined twelve probabilities, one for each seed count. These probabilities

24

^{518} are defined by age-specific germination rates and the experimental design, specifically the
^{519} timing of when seeds were unearthed.

^{520} Seeds from each experimental year and age were collected twice. First, the seed bags
^{521} were unearthed in the January as seeds were germinating. At this point, we counted intact
^{522} seeds and emerging seedlings and summed these to get a count of total seeds just before
^{523} germination. The seed bags were then returned to the ground from January to October. At
^{524} this point, the seed bags were unearthed again and removed from the field to get counts of
^{525} intact seeds.

^{526} Event histories 1-6 correspond to seed bags buried in experimental year 1; event histories
^{527} 7-10 correspond to seed bags buried in experimental year 2; event histories 11-12 correspond
^{528} to seed bags buried in experimental year 3. The probabilities are composed of germination
^{529} probabilities for seeds of age 1 ($\gamma_{..1}$), age 2 ($\gamma_{..2}$), and age 3 ($\gamma_{..3}$). The germination prob-
^{530} abilities are indexed by site j, experimental year k, and age l; for example, germination of
^{531} 1-year old seeds at site j from the experimental year k is given by γ_{jkl}. The compound event
^{532} histories are defined by:

$$
\begin{aligned}
\theta_{j1,1} &= 1 \\
\theta_{j1,2} &= (1 - \gamma_{j1,1}) \\
\theta_{j1,3} &= (1 - \gamma_{j1,1}) \\
\theta_{j1,4} &= (1 - \gamma_{j1,1}) \times (1 - \gamma_{j1,2}) \\
\theta_{j1,5} &= (1 - \gamma_{j1,1}) \times (1 - \gamma_{j1,2}) \\
\theta_{j1,6} &= (1 - \gamma_{j1,1}) \times (1 - \gamma_{j1,2}) \times (1 - \gamma_{j1,3}) \\
\theta_{j2,1} &= 1 \\
\theta_{j2,2} &= (1 - \gamma_{j2,1}) \\
\theta_{j2,3} &= (1 - \gamma_{j2,1}) \\
\theta_{j2,4} &= (1 - \gamma_{j2,1}) \times (1 - \gamma_{j2,2}) \\
\theta_{j3,1} &= 1 \\
\theta_{j3,2} &= (1 - \gamma_{j3,1})
\end{aligned}
\tag{9}
$$

^{533} These compound event histories ($\boldsymbol{\theta}$) link the germination model to the seed survival model.

They enter the likelihood statement in the definition of the deterministic process model. The probability of survival to time t_{ijkm} is now multiplied by the probability that a seed has not germinated by that time. That probability is defined by the compound event history leading up to time t_{ijkm}. The index m thus corresponds tracks the observation time and/or event history associated with that observation time.

6.2.4 Linking germination and seed survival

In the full model linking germination and seed survival data, counts of seeds are indexed by sites j, rounds of observation starting in a series of years k, seed ages l, and event histories m. The likelihood for the model is then

$$g(\eta_{ijk}, \theta_{jkm}, t_{ijkm}) = \theta_{jkm} \times \exp(-(\frac{t_{ijkm}}{\exp(-\frac{\eta_{ijk}}{\alpha_j})})^{\alpha_j})$$

$$[\boldsymbol{\mu}_{\mathrm{g}}, \boldsymbol{\sigma}_{\mathrm{g}}, \boldsymbol{\mu}_{\mathrm{g}}^{\mathrm{pop}}, \boldsymbol{\sigma}_{\mathrm{g}}^{\mathrm{pop}}, \boldsymbol{\eta}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\mu}^{\mathrm{pop}}, \boldsymbol{\sigma}^{\mathrm{pop}} | \mathbf{y}_{\mathrm{g}}, \mathbf{y}] \propto \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \left[\left[\prod_{m=1}^{M} \mathrm{binomial}(y_{ijkm} | n_{ijkm}, g(\theta_{jkm}, \eta_{ijk}, \alpha_j, t_{ijkm})) \right. \right.$$

$$\times \mathrm{normal}(\eta_{ijk} | \mu_{jk}, \sigma_{jk})$$

$$\times \mathrm{normal}(\mu_{jk} | \mu_j^{\mathrm{pop}}, \sigma_j^{\mathrm{pop}}) \mathrm{half\text{-}normal}(\sigma_{jk} | 0, 1)$$

$$\times \mathrm{normal}(\mu_j^{\mathrm{pop}} | 0, 1) \mathrm{half\text{-}normal}(\sigma_j^{\mathrm{pop}} | 0, 1)$$

$$\left. \times \mathrm{gamma}(\alpha_j | 2, 2) \right]$$

$$\left[\times \prod_{l=1}^{L} \mathrm{binomial}(y_{\mathrm{g},ijkl} | n_{\mathrm{g},ijkl}, \mathrm{logit}^{-1}(\alpha_{\mathrm{g},ijkl})) \right.$$

$$\times \mathrm{normal}(\alpha_{\mathrm{g},ijkl} | \mu_{\mathrm{g},jkl}, \sigma_{\mathrm{g},jkl})$$

$$\times \mathrm{normal}(\mu_{\mathrm{g},jkl} | \mu_{\mathrm{g},jl}^{\mathrm{pop}}, \sigma_{\mathrm{g},jl}^{\mathrm{pop}}) \mathrm{half\text{-}normal}(\sigma_{\mathrm{g},jkl} | 0, 1)$$

$$\left. \left. \times \mathrm{normal}(\mu_{\mathrm{g},jl}^{\mathrm{pop}} | 0, 1) \mathrm{half\text{-}normal}(\sigma_{\mathrm{g},jl}^{\mathrm{pop}} | 0, 1). \right] \right]$$

$$(10)$$

6.3 Estimating seed survival from reproduction to October

The model described so far uses data from the seed bag experiment to estimate conditional germination and seed survival through time. However, we can't use this data to estimate the probability of seeds surviving from reproduction to the time the seed bag experiment started in October (s_0). What we do to try to get this value is the following.

We use the counts of fruits per plant conducted in the permanent plots to get a total number of fruits per plot. We then multiply this count by the average number of seeds per fruit. We then have an estimate of the number of seeds produced per plot. We do this for aboveground data in 2007 and 2008. We link this data to the number of seedlings observed in the same plots in the following year (2008 and 2009, respectively). We thus take the total number of seeds produced in year t as the number of trials in a binomial experiment for which the outcome is the number of seedlings observed in year $t + 1$. The probability is the product of survival from reproduction to October (s_0, estimated here), survival from October to January, and germination. We have estimates of the latter two probabilities thanks to the seed bag experiments. We link the three components and estimate the remaining term s_0.

For the time being, we assume that the majority of seedlings in a plot emerge from seeds produced the previous year (read: germination of 1 year old seeds). With this assumption, we do not model germination from older seeds. We are also unable to assess the contribution of seeds from outside the plot to the total number of seeds available. Because the number of seedlings sometimes exceed the total number of seeds, we summed across transects to get counts for the number of seeds and observed seedlings.

In the this model, counts of seeds are indexed by sites j and years k. The likelihood for the model is then

$$[\boldsymbol{\mu}_{\mathrm{s}}, \boldsymbol{\sigma}_{\mathrm{s}}, \boldsymbol{\mu}_{\mathrm{s}}^{\mathrm{pop}}, \boldsymbol{\sigma}_{\mathrm{s}}^{\mathrm{pop}}|\mathbf{y}_{\mathrm{s}}] \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \mathrm{binomial}(y_{\mathrm{g},ijk}|n_{\mathrm{g},ijk}, \boldsymbol{\theta}\boldsymbol{\eta}\mathrm{logit}^{-1}(\alpha_{\mathrm{s},ijk}))$$

$$\times \mathrm{normal}(\alpha_{\mathrm{s},ijk}|\mu_{\mathrm{g},jk}, \sigma_{\mathrm{s},jk})$$

$$\times \mathrm{normal}(\mu_{\mathrm{s},jk}|\mu_{\mathrm{s},j}^{\mathrm{pop}}, \sigma_{\mathrm{s},j}^{\mathrm{pop}}) \tag{11}$$

$$\times \mathrm{half\text{-}normal}(\sigma_{\mathrm{s},jk}|0, 1)$$

$$\times \mathrm{normal}(\mu_{\mathrm{s},j}^{\mathrm{pop}}|0, 1)\mathrm{half\text{-}normal}(\sigma_{\mathrm{s},j}^{\mathrm{pop}}|0, 1).$$

This likelihood can be augmented to the one for the seed bag burial experiment.

6.4   Simulations

I conducted simulations to check whether the model formulation for the seed bag experiment above was able to recover parameters of known values. I generated data similar to what we observed in the experiment but, for the time, ignoring the hierarchical structure of the data).

I simulated data with known combinations of the shape and scale parameters for the Weibull, age-dependent germination probabilities, and aboveground plot dynamics. I checked for convergence and recovered the simulated parameters. I also performed graphical posterior predictive checks and calculated Bayesian p-values to assess how the model might perform when fit to real data.
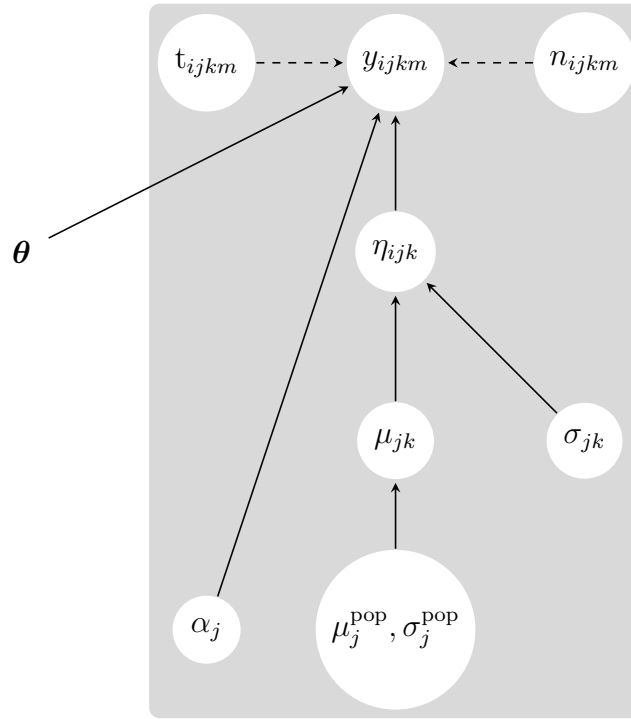
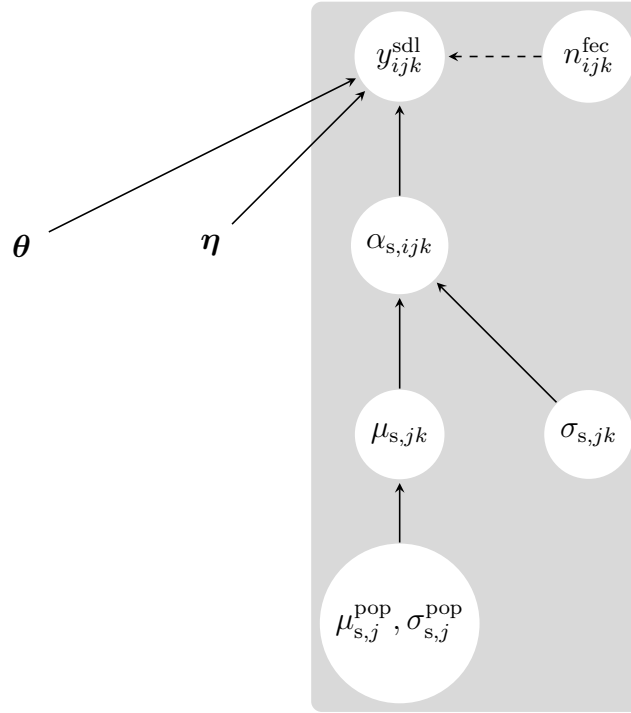## 6.5   Directed acyclic graphs

6.5.1   Germination
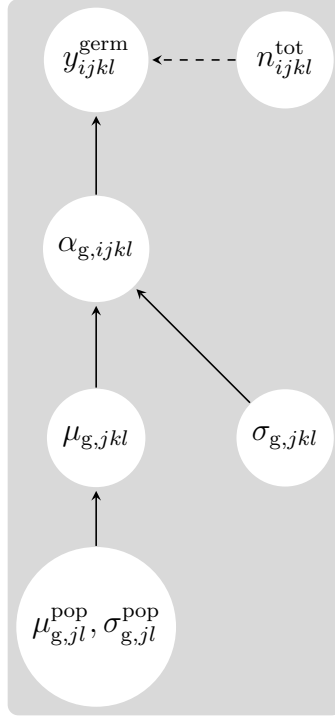
(a) Directed acyclic graph for the hierarchical model for conditional germination. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

6.5.2   Seed survival

(a) Directed acyclic graphs for the models of seed survival. The parameter $\boldsymbol{\theta}$ links the conditional germination model to the seed survival model. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

6.5.3 Seed survival from June to October

(a) Directed acyclic graphs for the models of seed survival from reproduction to October. The parameter $\boldsymbol{\theta}$ links the model to the conditional germination model. The parameter $\boldsymbol{\eta}$ links the model to the seed survival model. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

6.5.4 Full model: germination, seed survivorship, initial seed survival
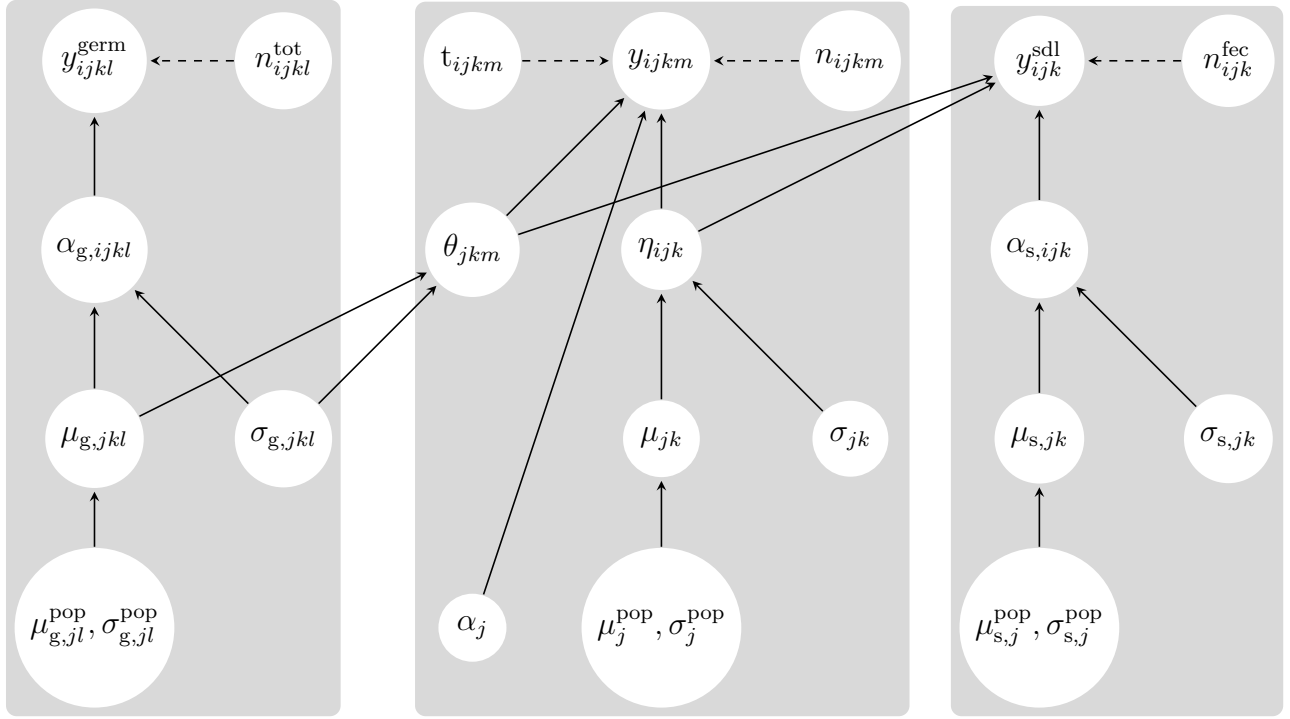
(a) Directed acyclic graphs for the full model. Models for each set of data are encapsulated in a gray box; links among the datasets are shown by arrows that cross over between boxes. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

7   Viability trials

7.1   Data

We collected the following data during the viability trials:

- n_{ijkl}^{germ} = observed count of seeds at the start of the germination trial for the ith bag, from the jth population, in the kth experimental year, for seeds of age l, assumed to be measured perfectly

- y_{ijk}^{germ} = observed count of germinated seedlings in the germination trial for the ith bag, from the jth population, in the kth year, for seeds of age l, assumed to be measured perfectly

- n_{ijk}^{viab} = observed count of seeds at the start of the viability trial for the ith bag, from the jth population, in the kth year, for seeds of age l, assumed to be measured perfectly

- y_{ijk}^{viab} = observed count of viable seedlings in the viability trial for the ith bag, from the jth population, in the kth year, for seeds of age l, assumed to be measured perfectly

7.2   Model

All data from viability trials is in the form of binomial trials: we have counts of seeds at the start and end of an experimental window of time. All models have the same structure for seeds in bag i in population j in experimental year k. If the number of seeds starting the trial (trials) is n_{ijk} and the number of seeds at the end of the trial (successes) is y_{ijk}, we write a model that has a population-level mean and year-level means drawn from the population-level distribution. Broadly, this is two-level hierarchical model with a population-level mean, and year-level means drawn from the population-level distribution. The probability of success for each bag is drawn from this year- and population-level distribution. The model uses a

33

binomial likelihood. The joint posterior for viability trials thus has the following general form:

$$[\boldsymbol{\alpha_G}, \boldsymbol{\mu_G}, \boldsymbol{\sigma_G}, \boldsymbol{\mu_G^{pop}}, \boldsymbol{\sigma_G^{pop}}|\boldsymbol{y^{tot}}] \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \text{binomial}(y_{ijk}^{germ}|n_{ijk}^{test_g}, \text{logit}^{-1}(\alpha_{G,ijk}))$$

$$\times \text{normal}(\alpha_{G,ijk}|\mu_{G,jk}, \sigma_{G,jk})$$

$$\times \text{normal}(\mu_{G,jk}|\mu_{G,j}^{pop}, \sigma_{G,j}^{pop}) \tag{12}$$

$$\times \text{half-normal}(\sigma_{G,jk}|0, 1)$$

$$\times \text{normal}(\mu_{G,j}^{pop}|0, 1000)\text{half-normal}(\sigma_{G,j}^{pop}|0, 1).$$

$$[\boldsymbol{\alpha_V}, \boldsymbol{\mu_V}, \boldsymbol{\sigma_V}, \boldsymbol{\mu_V^{pop}}, \boldsymbol{\sigma_V^{pop}}|\boldsymbol{y^{tot}}] \propto \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K} \text{binomial}(y_{ijk}^{viab}|n_{ijk}^{test_v}, \text{logit}^{-1}(\alpha_{V,ijk}))$$

$$\times \text{normal}(\alpha_{V,ijk}|\mu_{V,jk}, \sigma_{V,jk})$$

$$\times \text{normal}(\mu_{V,jk}|\mu_{V,j}^{pop}, \sigma_{V,j}^{pop}) \tag{13}$$

$$\times \text{half-normal}(\sigma_{V,jk}|0, 1)$$

$$\times \text{normal}(\mu_{V,j}^{pop}|0, 1000)\text{half-normal}(\sigma_{V,j}^{pop}|0, 1).$$

7.3 Directed acyclic graphs
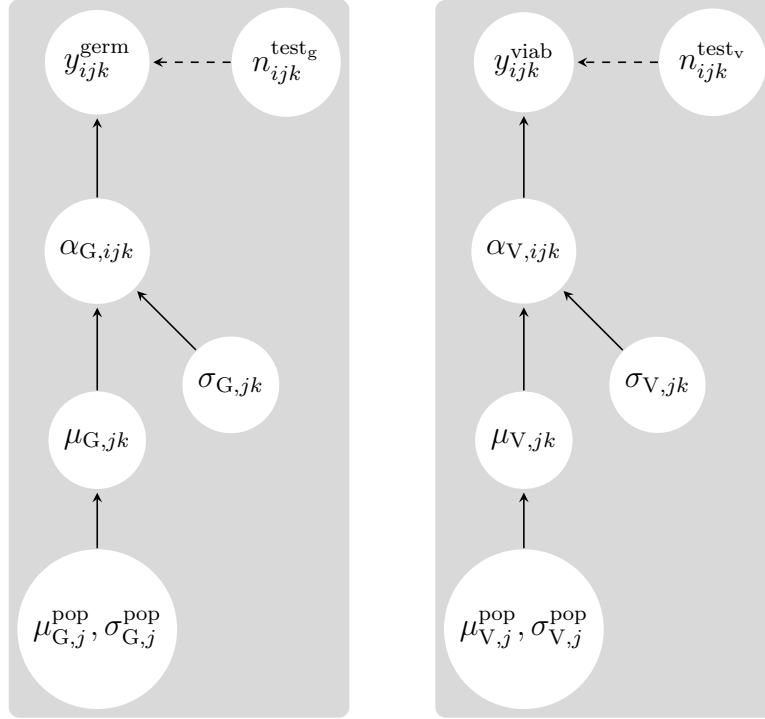
(a) Directed acyclic graphs for the hierarchical models for viability trials. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

8 Estimated parameters to model parameters

We now use the parameter estimates from Section 6 and 7 to obtain the parameters for the structured population model from Section 5. Below, we describe how we obtained population-level estimates for the parameters in Table 5.

First, we use the full model for germination, seed survival, and initial seed survival to obtain population-level estimates for the survival functions for seed persistence and age-specific germination. We use the posterior distributions for the population-level mean of the inverse-scale parameter (μ_j^{pop}) and the shape parameter (α_j) to derive (semantics: marginalize to obtain?) the population-level parameters for the Weibull survival function. We transform

the population-level posterior distributions for the survival probability from the latent scale to $[0, 1]$ by taking the inverse logit; this transforms the parameters into the probability of success.

We then use the posterior distributions for the population-level mean of the probability of germination on the latent scale ($\mu_{\text{g},jl}^{\text{pop}}$) to derive the age-specific germination probabilities. We transform the population-level posterior distributions for the germination probability from the latent scale to $[0, 1]$ by taking the inverse logit; this transforms the parameters into the probability of success.

We use the product of the Weibull survival function and the germination histories (as in Section 6) to obtain the survival functions associated with persistence of seeds in the soil seed bank (θ. in Table 4). At this point, we switch from continuous time to the discrete times in Table 4) by calculating the survival at those times.

We then estimate viability with the posterior distributions for the population-level mean of the probability of germination and viability from the viability trials. For estimates of viability in October, we transform the values from the latent scale ($\mu_{\text{G},j}^{\text{pop}}, \mu_{\text{V},j}^{\text{pop}}$) to the probability scale (θ_G, θ_V. We estimate the overall probability of viability, ν, as $\theta_g + \theta_v(1 - \theta_g)$. This weights the estimates relative to the probability of germination (eg. if no seeds germinate the estimate of viability will mostly come from the viability test). For estimates of viability in January, we interpolate between estimates of viability on the latent scale before transforming interpolated values to the probability scale. This allows us to retain the uncertainty associated with our estimates of viability. We transform the population-level posterior distributions for the overall probability of viability from the latent scale to $[0, 1]$ by taking the inverse logit; this transforms the parameters into the probability of success.

With all of these pieces in hand, we calculate values of ϕ, which are the probability of seed persistence and viability. We then estimate the probability of germination conditional on viability as in Table 5.
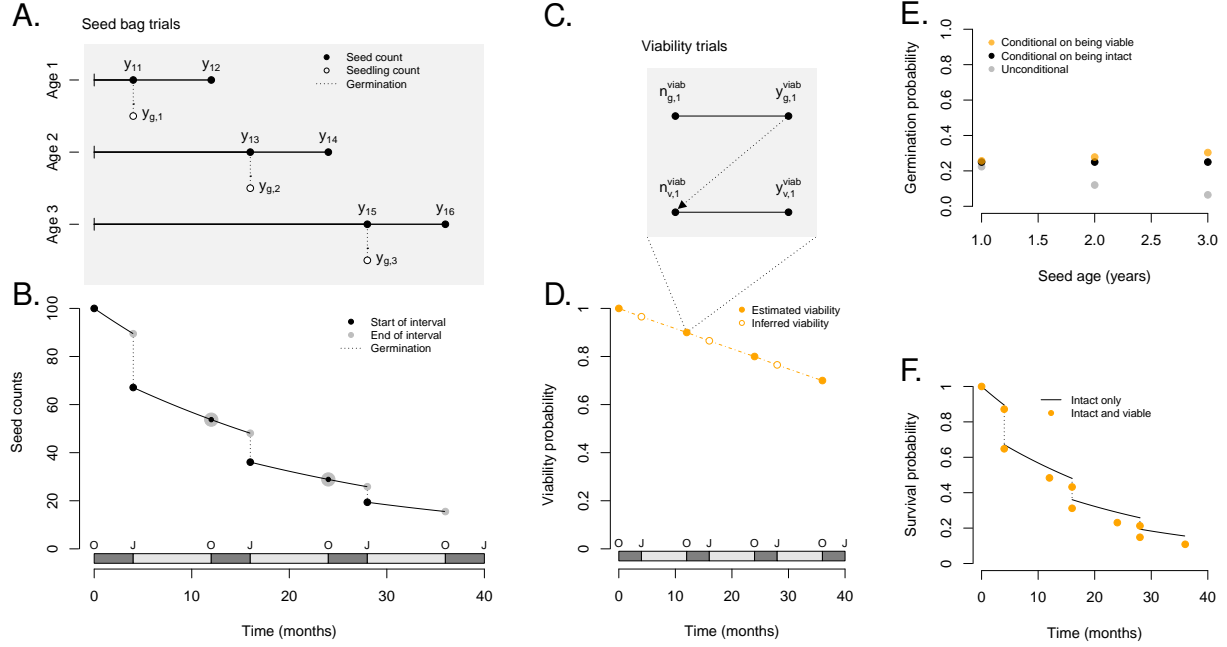
Figures

Figure 6: Summary of the seed bag burial experiments and viability trials. <mark>Figure will be labeled as (A: seed bag trials , B: viability trials, C: germination probability, D: survival probability.</mark> . (A) The gray panel contains a graphical representation of the seed bag trials. Seeds were buried at the start of each experiment (100 seeds in month 0). Seed bags were unearthed and intact seeds ($y_{..}$) and germinants ($y_{g,.}$ counted. The graph below the panel shows a hypothetical survival function associated with persistence of seeds in the soil seed bank. (B) The gray panel contains a graphical representation of the viability trials. Seeds were tested in two rounds; germination trials were performed and then some or all of the ungerminated seeds were tested for viability. The graph below the panel shows hypothetical data from a series of viability trials and the interpolated, inferred viabilities at times when viability was unobserved. (C) Age-specific germination probably is summarized in three ways. (D) The graph shows the survival function for persistence of seeds in the soil seed bank (black line) and the estimated discrete survival probabilities for persistence and viability of seeds (orange points).
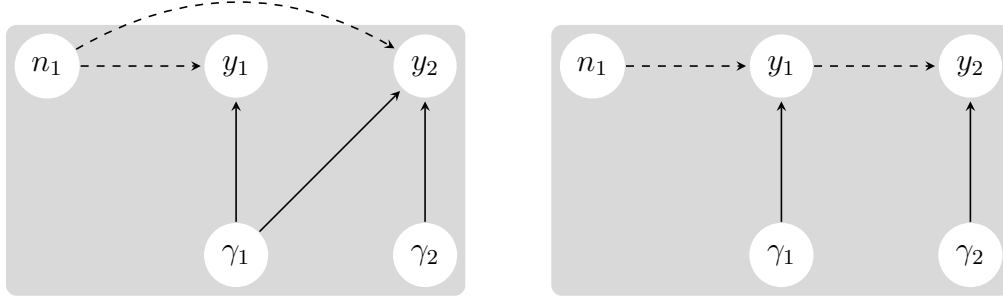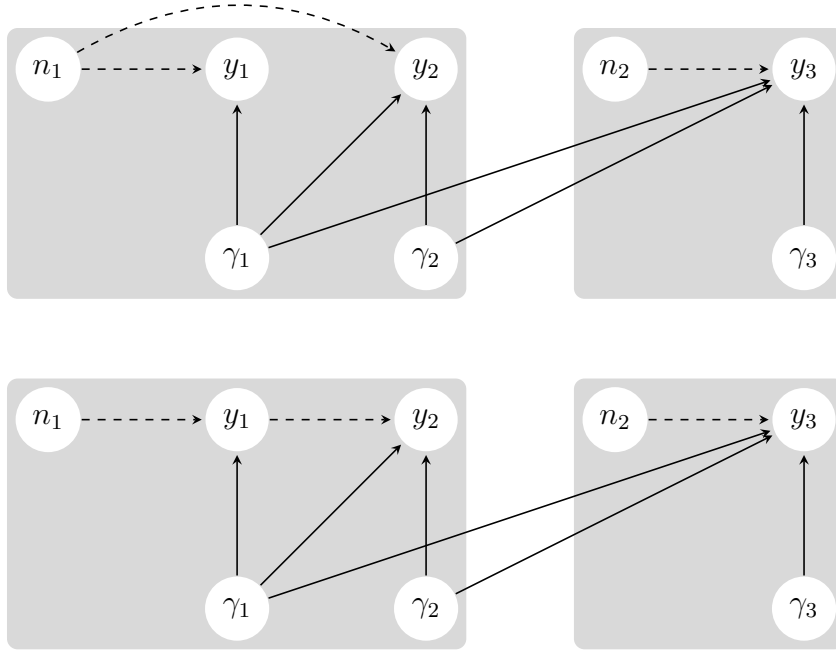
37

(a) Directed acyclic graphs for the hierarchical models for seed bag rates. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

$$[\gamma_1, \gamma_2 | \boldsymbol{y_1}, \boldsymbol{y_2}] \propto \text{binomial}(y_1 | n_1, \text{logit}^{-1}(\alpha_1))$$

$$\times \text{binomial}(y_2 | n_1, \text{logit}^{-1}(\alpha_1) \times \text{logit}^{-1}(\alpha_2))$$

$$\times \text{normal}(\alpha_1 | \mu_1, \sigma_1) \text{normal}(\alpha_2 | \mu_2, \sigma_2) \tag{14}$$

$$\times \text{normal}(\mu_1 | 0, 1000) \text{half-normal}(\sigma_1 | 0, 1)$$

$$\times \text{normal}(\mu_2 | 0, 1000) \text{half-normal}(\sigma_2 | 0, 1).$$

$$[\gamma_1, \gamma_2 | \boldsymbol{y_1}, \boldsymbol{y_2}] \propto \text{binomial}(y_1 | n_1, \text{logit}^{-1}(\alpha_1))$$

$$\times \text{binomial}(y_2 | y_1, \text{logit}^{-1}(\alpha_2))$$

$$\times \text{normal}(\alpha_1 | \mu_1, \sigma_1) \text{normal}(\alpha_2 | \mu_2, \sigma_2) \tag{15}$$

$$\times \text{normal}(\mu_1 | 0, 1000) \text{half-normal}(\sigma_1 | 0, 1)$$

$$\times \text{normal}(\mu_2 | 0, 1000) \text{half-normal}(\sigma_2 | 0, 1).$$

Identifiability across years

(a) Directed acyclic graphs for the hierarchical models for seed bag rates. Solid arrows depict the relationships among random variables, and dashed arrows depict the deterministic relationships.

$$
\begin{aligned}
[\gamma_1, \gamma_2 | \boldsymbol{y_1}, \boldsymbol{y_2}] \propto\ & \text{binomial}(y_1 | n_1, \text{logit}^{-1}(\alpha_1)) \\
& \times \text{binomial}(y_2 | n_1, \text{logit}^{-1}(\alpha_1) \times \text{logit}^{-1}(\alpha_2)) \\
& \times \text{normal}(\alpha_1 | \mu_1, \sigma_1)\text{normal}(\alpha_2 | \mu_2, \sigma_2) \\
& \times \text{normal}(\mu_1 | 0, 1000)\text{half-normal}(\sigma_1 | 0, 1) \\
& \times \text{normal}(\mu_2 | 0, 1000)\text{half-normal}(\sigma_2 | 0, 1).
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
[\gamma_1, \gamma_2 | \boldsymbol{y_1}, \boldsymbol{y_2}] \propto\ & \text{binomial}(y_1 | n_1, \text{logit}^{-1}(\alpha_1)) \\
& \times \text{binomial}(y_2 | y_1, \text{logit}^{-1}(\alpha_2)) \\
& \times \text{normal}(\alpha_1 | \mu_1, \sigma_1)\text{normal}(\alpha_2 | \mu_2, \sigma_2) \\
& \times \text{normal}(\mu_1 | 0, 1000)\text{half-normal}(\sigma_1 | 0, 1) \\
& \times \text{normal}(\mu_2 | 0, 1000)\text{half-normal}(\sigma_2 | 0, 1).
\end{aligned}
\tag{17}
$$

In constructing our survival function, we have two types of data for each experimental year and age. Counts from January represent totals from prior to the germination of seeds of that age. The survival function of seeds at time T, $S_{ya}^1(T)$, for these counts in each experimental year y is described by

$$S_a^{\text{jan}} = \prod_{j=1}^{a}(1 - \gamma_{yj})^{I(j=j-1)} \tag{18}$$

Counts from October represent totals from after the germination of seeds of that age. The survival function of seeds at time T, $S_{ya}^2(T)$, for these counts in each experimental year y is described by

$$S_a^{\text{oct}} = \prod_{j=1}^{a}(1 - \gamma_{yj}) \tag{19}$$

The functions S^{jan} and S^{oct} describe the probability that a seed with a particular event (germination) history survives to January and October, respectively.

To include both types of data in the seed bag experiment model, I combined the two functions and introduced an additional index for the type of data the function describes. The index $d = 1, 2$ now represents whether the data are counts from January ($d = 1$) or October ($d = 2$)

$$S_{da} = \prod_{j=1}^{a}\prod_{i=1}^{d}(1 - \gamma_{yj})^{I(d=2)} \tag{20}$$

The probability described below could probably be expressed compactly with indicator functions but I'm running into blocks on getting this to work. Below, $I(\cdot)$ is an indicator function that takes on the value of 1 when the statement in parentheses is true and the value

₆₂₄ of 0 when the statement in parentheses is false.

₆₂₅ The first indicator function indicates whether we are observing the total number of
₆₂₆ seedlings in January (before germination).

627 $$\gamma_{jkl} = \left(1 - \gamma_{1_{kl}}\right)^{(1-I(j=1 \wedge l=1))} \times \left(1 - \gamma_{2_{kl}}\right)^{(1-I(j=1 \wedge l=2))} \times \left(1 - \gamma_{3_{kl}}\right)^{(1-I(j=1 \wedge l=3))}$$

The models below represent the joint likelihood for data from the seed bag experiments. All data from seed bags and viability trials is in the form of binomial trials: we have counts of seeds at the start and end of an experimental window of time. All models for the parameters $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$ have the same structure for seeds in bag i in year j in population k. If the number of seeds starting the trial (trials) is n_{ijk} and the number of seeds at the end of the trial (successes) is y_{ijk}, we write a model that has a population-level mean and year-level means drawn from the population-level distribution. The probability of success for each bag is drawn from this year- and population-level distribution:

I compared convergence diagnostics (R-hat, effective sample size) for centered and non-centered parameterizations of the model. Here, I use the centered parameterization because this led to improved convergence. In each model, we obtain the population-level posterior distribution probability of success (the γs) by marginalizing across years and taking the inverse logit.

The seed bank experiments might used to obtain age-specific survival and recruitment terms. So the terms could be survival to month 3 (equal to age-specific survival at midpoint), germination at month 3, survival from month 3 to month 12 (equal to age-specific survival at midpoint), germination at month 15, survival from month 15 to 24, germination at month 27, survival from month 27 to 36. There are some errors in the data still, need to look at those. For germination there is no clear pattern to germination, either for individual sites or across sites. There are monotonic increases, monotonic decreases, and nonmonotonic patterns. Might also help to look at viability across age.

The seed pot experiments seem like they would be amenable to the kind of functions fit in the paper by Rees and Long. In those cases there are only 3 time points which would limit the application of the empirical seedling recruitment curves. But perhaps overlaying them per population would allow us to figure out whether there is consistency across years or whether particular cohorts follow different curves, and how this varies in space?

I will estimate survival/mortality rates at 4 time points (0-3 months, 3-12 months, 15-24 months, 27-36 months), and germination at 3 time points (3 months, 15 months, 27 months). This will allow me to figure out what kind of model (cf Rees and Long) might be most appropriate for the data, especially for the emergence data from the seed pot trials.

Table 6: Summary of models in Rees and Long (1993).

Model	Description
Exponential	...
Compound exponential	...
Weibull	...
Log logistic	...

We start with

$$y_{it} \sim \text{binomial}(n_i, \gamma)$$

$$\gamma \sim \text{beta}(1, 1) \tag{21}$$

to say that the observations of the number of seeds counted in bag i at time t are represented as y_{it} drawn from a binomial distribution. In this case, n_i is the number of trials, the number of seeds starting the experiment; y_{it} is the number of successes, the number of seeds remaining. The γ is the probability of success on a single trial.

$$[\gamma | \boldsymbol{y}, \boldsymbol{n}] \propto \text{binomial}(y_{it} | n_i, \gamma) \text{beta}(\gamma | 1, 1) \tag{22}$$

We then have sampling variability that is implicit in the binomial. There are thus two sources of uncertainty. Uncertainty arising from sampling and uncertainty arising because of . . . .

The case where each bag has its own mean survival at each time, acknowledging variation

45

among bags:

$$y_{it} \sim \text{binomial}(n_i, \gamma_{it})$$

$$\gamma_{it} \sim \text{beta}(\alpha, \beta)$$

$$(23)$$

Giving the following :

$$[\boldsymbol{\gamma}, \alpha, \beta | \boldsymbol{y}, \boldsymbol{n}] \propto \text{binomial}(y_{it} | n_i, \gamma_{it}) \text{beta}(\gamma_{it} | \alpha, \beta)$$

$$\text{gamma}(\alpha | 0.001, 0.001) \text{gamma}(\beta | 0.001, 0.001)$$

$$(24)$$

We now want to model the process "change in being intact with time" as

$$g(k, t_i) = \exp(-k t_i)$$

$$(25)$$

This deterministic model represents the average proportion of seeds that are expected to be intact as a function of time under a negative exponential process. The function $g(k, t_i)$ is the overall mean probability of being intact at time t_i. This is the mean of the beta distribution; the variance is the variation in probability of being intact at time t_i that arises from differences among time. The uncertainty that arises from sampling - which we can estimate because of replication - is distinct from this process variance. I think the process variance is all effects that create variation beyond the seed age, as represented with a negative exponential function.

The next problem is that the process is not simply one of decay, decomposition, or mortality. Instead, there are annual events interspersed into this, namely germination. In this way I think the situation resembles the complement of case 2 in Olson (1963). Modeling

germination and survival jointly would account for the full data. Here's one idea:

$$h_1(k, t_1) = \exp(-kt_1)$$

$$h_2(k, t_1) = g_1 \exp(-kt_1)$$

$$h_3(k, t_2) = (1 - g_1) \exp(-kt_2)$$

$$h_4(k, t_3) = (1 - g_1) \exp(-kt_3)$$

$$h_5(k, t_3) = (1 - g_1)(g_2) \exp(-kt_3) \tag{26}$$

$$h_6(k, t_4) = (1 - g_1)(1 - g_2) \exp(-kt_4)$$

$$h_7(k, t_5) = (1 - g_1)(1 - g_2) \exp(-kt_5)$$

$$h_8(k, t_5) = (1 - g_1)(1 - g_2)g_3 \exp(-kt_5)$$

$$h_9(k, t_6) = (1 - g_1)(1 - g_2)(1 - g_3) \exp(-kt_6)$$

Perhaps it would be productive to break the probability into two processes, one accounting for mortality and another accounting for decay. The one accounting for mortality would follow one of the following processes. First, we consider a negative exponential mortality trajectory.

$$g(k, t_i) = \exp(-kt_i) \tag{27}$$

Then, we consider a continuous exponential mortality trajectory:

$$g(a, b, t_i) = \frac{1}{(1 + bt_i)^a} \tag{28}$$

Then, we consider a mortality trajectory for Weibull residence times:

$$g(\alpha, \beta, t_i) = \exp -(\frac{t_i}{\beta})^\alpha \tag{29}$$

The mortality process would be multiplied with the process "change in removal from population with age due to germination", represented as

$$
\begin{aligned}
h_1(\boldsymbol{g}) &= 1 \\
h_2(\boldsymbol{g}) &= g_1 \\
h_3(\boldsymbol{g}) &= (1 - g_1) \\
h_4(\boldsymbol{g}) &= (1 - g_1) \\
h_5(\boldsymbol{g}) &= (1 - g_1)g_2 \\
h_6(\boldsymbol{g}) &= (1 - g_1)(1 - g_2) \\
h_7(\boldsymbol{g}) &= (1 - g_1)(1 - g_2) \\
h_8(\boldsymbol{g}) &= (1 - g_1)(1 - g_2)g_3 \\
h_9(\boldsymbol{g}) &= (1 - g_1)(1 - g_2)(1 - g_3)
\end{aligned}
\tag{30}
$$

In turn we would consider models that combine age-dependent and -independent germination functions and mortality functions.

The following equations correspond to the full conditional distributions for models at the first time point. Subsequent models incorporate germination in the function $m(\ldots, \sigma^2)$. Specifically, the mean is the product of the mortality up to time t_i, represented by $g(\ldots)$, germination up to time t_i, represented by $h(\ldots)$. We will use the information from the seed bag experiments to determine whether there is an appropriate mortality process with which to model the recruitment data from the seed pots. What I would like to do with this is get a mortality and germination process that's appropriate for these sites and use the posterior for the parameters in that process as informed priors to model the seed recruitment from the seed pots.

Finally, we also consider the model proposed in Gremer and Venable (2014), namely one

48

in which seed mortality is seasonal and age-dependent.

$$g(k, t_i) = \exp(-kt_i)$$

$$[k, \alpha, \beta | \boldsymbol{y}, \boldsymbol{n}] \propto \text{binomial}(y_{it}|n_i, \gamma_{it})\text{beta}(\gamma_{it}|m(g(k, t_i), \sigma^2))$$

$$\times \text{gamma}(k|0.001, 0.001)$$

$$\times \text{inverse gamma}(\sigma^2|0.001, 0.001)$$

$$(31)$$

$$g(a, b, t_i) = \frac{1}{(1 + bt_i)^a}$$

$$[k, \alpha, \beta | \boldsymbol{y}, \boldsymbol{n}] \propto \text{binomial}(y_{it}|n_i, \gamma_{it})\text{beta}(\gamma_{it}|m(g(a, b, t_i), \sigma^2))$$

$$\times \text{gamma}(a|0.001, 0.001)\text{gamma}(b|0.001, 0.001)$$

$$\times \text{inverse gamma}(\sigma^2|0.001, 0.001)$$

$$(32)$$

$$g(\alpha, \beta, t_i) = \exp -(\frac{t_i}{\beta})^\alpha$$

$$[k, \alpha, \beta | \boldsymbol{y}, \boldsymbol{n}] \propto \text{binomial}(y_{it}|n_i, \gamma_{it})\text{beta}(\gamma_{it}|m(g(\alpha, \beta, t_i), \sigma^2))$$

$$\times \text{gamma}(\alpha|0.001, 0.001) \times \text{gamma}(\beta|0.001, 0.001)$$

$$\times \text{inverse gamma}(\sigma^2|0.001, 0.001)$$

$$(33)$$

Modeling seed survival

683 We chose to work with a Weibull residence time model, which represents the seed pool as

684 a distribution of survival times. For background on this see Pinder et al. (1978); Cornwell

685 and Weedon (2014); Dahlgren et al. 2016. Parameter alpha controls the shape of the

686 decomposition trajectory and beta the rate of decomposition. This model can reduce to the

687 exponential model when alpha is 1. If alpha is less than 1, the decomposition rate decreases

688 through time. If alpha is greater than 1, the decomposition rate increases through time. (see

689 Pinder et al. 1978 for another description).

Consider a survivorship trajectory for Weibull residence times:

$$g(\alpha, \beta, t_i) = \exp{-(\frac{t_i}{\beta})^\alpha} \tag{34}$$

The mortality process would be multiplied with the process "change in removal from population with age due to germination", represented as

$$
\begin{aligned}
h_1(\boldsymbol{g}) &= 1 \\
h_3(\boldsymbol{g}) &= (1 - g_1) \\
h_4(\boldsymbol{g}) &= (1 - g_1) \\
h_6(\boldsymbol{g}) &= (1 - g_1)(1 - g_2) \\
h_7(\boldsymbol{g}) &= (1 - g_1)(1 - g_2) \\
h_9(\boldsymbol{g}) &= (1 - g_1)(1 - g_2)(1 - g_3)
\end{aligned}
\tag{35}
$$

I compactly represent this history as the product of germination history and survivorship. I describe germination history as the product of a matrix of binary values and a vector of corresponding probabilities of not germinating. In the matrix of binary values, each row corresponds to a sampling instance in the seed bag experiment. Columns 1-3 correspond to

whether the seeds of age 1-3 had germinated at the respective sampling event. The vector describes the probability that seeds of age 1-3 did not germinate. The vector is in turn multiplied with the survival time function. These terms combine the loss of seeds from germination and the loss of seeds from decay/mortality.

$$h(\alpha, \beta, t_i) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 - g_1 & 1 - g_2 & 1 - g_3 \end{bmatrix}^\top \exp{-(\frac{t_i}{\beta})^\alpha} \tag{36}$$

690   Because we model germination as conditionally dependent on the number of intact seeds

691   at the time of germination, we estimate the germination rate independently. We combine

692   that information with the information from the survival time model. We use the survival

693   function associated with a Weibull distribution as the deterministic function for how seed

694   survivorship changes over time. The Weibull has a shape and scale parameter; we estimate

695   a shape parameter for each population. See Smits 2015 for logic from survival analysis. This

696   is equivalent to saying that we are assuming the change in survivorship is a population-level

697   property but that the scale varies from year to year within each population.

698   Because germination terms are present in both the survival and germination models,

699   data on both seed survivorship and germination inform estimates of germination. Including

700   the germination terms in the survivorship model means that the expected number of seeds

701   surviving to a given time is the product of a survival function and germination. With low

702   germination proportions, the dominant process is mortality. As germination increases, the

Pinder et al. (1978) also give the mean longevity calculated from a Weibull:

$$\text{scale} \times \Gamma(1 + (\frac{1}{\text{shape}})) \tag{37}$$

Which seems like an interesting extension of the data on hand. If we don't assume an exponential decay model of seed survival then the value of both the shape and scale parameters are relevant. So populations with an, on average, higher longevity would correspond to being a safer seed bank.

References

707 Doak, D. F., D. Thomson, and E. S. Jules. 2002. Population Viability Analysis for Plants:
708 Understanding the Demographic Consequences of Seed Banks for Population Health. In
709 Population viability analysis, pages 312–337.

710 Nguyen, V., Y. M. Buckley, R. Salguero-Gómez, and G. M. Wardle. 2019. Consequences of
711 neglecting cryptic life stages from demographic models. Ecological Modelling, **408**:108723.

712 Paniw, M., P. F. Quintana-Ascencio, F. Ojeda, and R. Salguero-Gómez. 2017. Accounting for
713 uncertainty in dormant life stages in stochastic demographic models. Oikos, **126**:900–909.
714 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/oik.03696.

715 Peterson, M. L., W. Morris, C. Linares, and D. Doak. 2019. Improv-
716 ing structured population models with more realistic representations of non-
717 normal growth. Methods in Ecology and Evolution, **10**:1431–1444. _eprint:
718 https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13240.