Hierarchical models for hard-to-see stages: estimating survivorship and germination in soil seed banks

Gregor Siegmund and Monica Geber
Department of Ecology and Evolutionary Biology, Cornell University

Slide 1

Transcript: Hello! I'll be talking about *Hierarchical models for hard-to-see stages: estimating survivorship and germination in soil seed banks*. My name is Gregor and I am a PhD student with Monica Geber. I'm excited to share the work that we've been doing motivated by long-term demographic data from the species *Clarkia xantiana.* Because this is entirely online, we won't have the opportunity to talk in person but I encourage you to reach out to me over email!

Slide 2

Transcript: Soil seed banks are critical for plant life history strategies that rely on bet hedging and predictive germination to persist in variable environments. Plant ecologists who want to incorporate seed banks into population models run into challenges – seeds are far more difficult to follow than aboveground plants. Although I've often heard that it's easier to collect demographic data on plants than animals because they don't move, I don't think that statement accounts for seed banks. In this talk, we present a modeling approach to estimate seed mortality and germination using observations from experiments with the seed bank.

Slide 3

Transcript: Here, I'm showing you a caricature of the demography modeling workflow in ecology. Previous research focused on methods to include seed banks in demographic studies has addressed the 3 points I've outlined here. First, ecologists have developed a variety of ways to quantify seed fates. A relatively popular method among population ecologists has been to construct experimental seed banks using bags or cages. Second, recent work has addressed how to propagate uncertainty about seed fates through to estimates of population growth rate. Third, studies have also conducted retrospective assessments to see how sensitive population growth is to changes in parameters or matrix structure.

Slide 4

Transcript: We're aiming to complement existing work by focusing on the part of the workflow that has received comparatively less attention. So here, we're going develop a deterministic model for observations from the seed bank, use a Bayesian model to estimate its parameters, and then see how we can apply model checking and selection.

Slide 5

Transcript: Seed banks are complex and most of the in- and out-flows are difficult or impossible to observe.

Slide 6

Transcript: For the purposes of demographic models, mortality and germination are the key processes. That's because seeds leave the seed bank either by germinating or by dying. The seeds that don't do either persist in the seed bank.

Slide 7

Transcript: Seed bag burial experiments are one of the ways that ecologists use to estimate seed mortality and germination. Of course there are others but we'll focus on this one because it's often possible to recover both intact, ungerminated seeds and seedlings from the buried bags. I'm showing you a sketch of the design for one of these experiments on the left, and the resulting data on the right.

Slide 8

Transcript: I've given you a general overview of the seed bank and experiments used to understand it. I'll now describe the deterministic model that I'm using to describe persistence in the seed bank. There's discrete and continuous versions of it but basically it takes the product of a seed's germination history and mortality history. This is because seeds that remain in the seed bank don't germinate or die.

Slide 9

Transcript: When we think about seed persistence as the product of not germinating and not dying, we can start to describe the age structure of the seed bank as a function of germination and mortality risk. If these don't vary over time, with seed age, we think of the seed bank as unstructured. On the other hand, if mortality risk increases with time then we introduce age structure into the seed bank. On this slide, I'm illustrating how these risks – also known as hazards – determine seed bank structure. The key idea is that we can use concepts from survival analysis to describe seed fates.

Slide 10

Transcript: Once we have a good understanding of the observations and plausible deterministic model for the process that generates those observations, we can write a statistical model. I've written a Bayesian model because it neatly shows how the deterministic model determines the probability of persistence over time, and allows to simultaneously fit seed and seedling counts. The model I'm showing you is just one possible version. For example, the equations here assume a constant risk of mortality. If we want to relax this assumption, we can use a different survival function such as the Weibull or the gamma.

Slide 11

Transcript: Next, we conducted a simulation study to validate our modeling approach. We simulated observations for a seed bag burial experiment. In those simulations, we varied four factors related to magnitude of germination and mortality, the study design, and the age structure of the seed bank. We fit several models to our simulated observations, and then evaluated the models based on a few different aspects. I'll use one slide each to give illustrations of how we assessed the models these categories.

Slide 12

Transcript: First, we used the same model to simulate and fit observations. For example, here we simulated observations from a seed bank without age structure and were then able to recover the germination and mortality probabilities used in the simulation when we fit a model that also did not assume age structure in mortality.

Slide 13

Transcript: Second, we simulated observations for seed banks with a range of age structures. To do this, we defined mortality by a Weibull which allowed us to vary the risk of mortality over time by modifying the shape parameter. We were interested in what would happen when we fit a model without age structure to observations generated by an age-structured process. To do this, we fit a model with a constant mortality hazard – basically we didn't allow the risk of mortality to vary over time. What the figure here illustrates is that the more mismatched the model is to the data generating process, the greater the bias in the parameter estimates. A value of 1 on the x-axis corresponds to an unstructured seed bank. As you move away from that the seed bank becomes more structured. The y-axis shows the error – essentially, the difference between the estimated value and the true value used in the simulation. Here, a value of 0 corresponds to no error. You can see that the error in estimates increases as seed bank structure increases.

Slide 14

Transcript: Finally, I'll briefly touch on what we tried in terms of model checks and selection. Field studies won't have the luxury of knowing what the data generating process is like – we won't know if it's an unstructured seed bank or not. So it would be great to have a way to decide, based on the data, whether we need to include age structure when we estimate parameters and when we construct the matrix model. The approach that worked reasonably well was to fit models with different structures to the same data and then compare them using cross-validation, which is a Bayesian model selection tool. You can see in the figure here that the model that imposes a constant mortality is not favored as the age structure becomes more severe, that is as you move away from 1 on the x-axis. As you move away from 1 on the x-axis, the negative exponential model, which has that constant risk of mortality, decreases in performance.

Slide 15

Transcript: To wrap up, I'll return to this slide that shows the parts of the modeling workflow that I discussed today. If you're interested, we have a draft of this work and encourage you to reach out to chat about it!

Slide 15

Transcript: I'd like to thank several people who have in one way or another prompted me to dig deeper into the questions I talked about today. I'd also like to thank the funding sources that have supported my graduate work, as well as the NSF grants that made it possible for Monica and her collaborators to collect the demographic observations that inspired this talk. Thanks very much for your time!