Edit roadmap

From our meeting last week, we discussed how the decision I made about what to include in the main text versus the supplement were, on the whole, reasonable. Some of the challenges that we discussed including how to frame the manuscript considering the results, how to communicate the methods, and how to introduce the idea of bet hedging and focus the paper.

I attempted to address all the comments you made on the draft by responding to them in this document. In doing so, I took account of what you wrote as well as the conversation that we had about the draft. I expect to work on incorporating all these changes into the manuscript.

In the **Introduction**, I will better connect theory about bet hedging to seed banks by explaining how bet hedging is thought to work. I will try to do this by describing the process rather than leaning on terminology to communicate with readers. Specifically, I will try to write about how variability in fitness interacts can favor the evolution of delayed germination and seed banks in the long term versus the short term. I will also better establish the conceptual/logical links underlying our decision to explore bet hedging in *Clarkia*. Specifically, I will establish the hypothesized relationships between seasonal precipitation, reproductive success, and seed bank rates. Finally, I will work to better integrate the questions I ask in the final paragraph throughout the introduction.

In the **Methods**, I will strengthen the connection between the questions/goals of the study, the data, the statistical methods, and the analysis. I will try to do this by foregrounding the connection at the start of the Methods, and referring readers back to these ideas throughout the Methods. I will also reduce the number of symbols and terms I use; I will make a table of symbols for readers to reference in the Supplement and I will emphasize germination and survival as they connect to bet hedging theory. I will provide a concrete example of the modelling process for seedling survival to fruiting and refer readers to the appendix for similar details on other vital rates. Finally, I will revise the section on estimating the belowground vital rates in order to make it clear to readers how the seed bag experiment is used.

Finally, we discussed three topics that I think are appropriate for the **Discussion**. First, I will explicitly discuss this study in context of complementary hypotheses for the presence of germination delays and seed banks. In doing so, I will discuss how I would test whether density-dependence or predictive germination contribute to the presence of a seed bank. Second, I will discuss the nature of environmental variability in the study populations, and how/whether the relationship between winter and spring precipitation & temperature could interact to provide a cue for predictive germination. Third, I will discuss the potential for genotype by environment interactions. For this, I will reference the variable relationships in how well precipitation predicts reproductive success, and place the study in the context of previous work with *Clarkia*.

I cite the following references in my response to comments:

Clauss, M. J., & Venable, D. L. (2000). Seed Germination in Desert Annuals: An Empirical Test of Adaptive Bet Hedging. *The American Naturalist*, *155*(2), 168–186.

Gremer, J. R., Kimball, S., & Venable, D. L. (2016). Within-and among-year germination in Sonoran Desert winter annuals: Bet hedging and predictive germination in a variable environment. *Ecology Letters*, *19*(10), 1209–1218.

Elderd, B. D., & Miller, T. E. X. (2016). Quantifying demographic uncertainty: Bayesian methods for integral projection models. *Ecological Monographs*, *86*(1), 125–144.

Ellner, S. (1985). ESS germination strategies in randomly varying environments. II. Reciprocal Yield-Law models. *Theoretical Population Biology*, *28*(1), 80–116.

Evans, M. E. K., Ferrière, R., Kane, M. J., & Venable, D. L. (2007). Bet Hedging via Seed Banking in Desert Evening Primroses (*Oenothera*, Onagraceae): Demographic Evidence from Natural Populations. *The American Naturalist*, *169*(2), 184–194.

Gremer, J. R., & Venable, D. L. (2014). Bet hedging in desert winter annual plants: Optimal germination strategies in a variable environment. *Ecology Letters*, *17*(3), 380–387.

Gremer, J. R., Kimball, S., & Venable, D. L. (2016). Within-and among-year germination in Sonoran Desert winter annuals: Bet hedging and predictive germination in a variable environment. *Ecology Letters*, *19*(10), 1209–1218.

Metcalf, C. J. E., Ellner, S. P., Childs, D. Z., Salguero‐Gómez, R., Merow, C., McMahon, S. M., Jongejans, E., & Rees, M. (2015). Statistical modelling of annual variation for inference on stochastic population dynamics using Integral Projection Models. *Methods in Ecology and Evolution*, *6*(9), 1007–1017.

Philippi, T. (1993). Bet-Hedging Germination of Desert Annuals: Variation Among Populations and Maternal Effects in *Lepidium lasiocarpum*. *The American Naturalist*, *142*(3), 488–507.

Main text

Introduction

**Line 6:** Maybe replace temporal with inter-annual. Temporal doesn’t specify a scale of variation. It could be decadal or within year variation. Whereas inter-annual is more specific.

GS: Change *temporal* to *interannual*

**Line 6:** Evolutionary ecologists have classically theorized seed banks as a bet-hedging strategy that maximizes geometric mean fitness across years. Add ‘across years’.

GS: add *across years* to sentence

**Line 13-14**: I think you need to expand on this first sentence. Imagine that you are a reader who is coming to this paper uninformed, both about the specifics of this paper and about bet hedging. Such a reader might not know why geometric mean fitness is even used. So, it would be useful to explain how seed banks (or more specifically seed dormancy) protect against failure – i.e. by virtue of maintaining some dormant seed, a parental plant guards against total reproductive failure should all of its seed germinate in a year that is unfavorable for survival and reproduction. Then, how having dormant seed can reduce mean arithmetic mean fitness but also reduce variance in fitness (isn’t it reducing the variance in fitness across years not the variance in the geometric mean fitness).

GS: Revise introduction to better connect seed banks to theory for readers new to the idea. I will focus on explaining how a strategy that reduces the arithmetic mean fitness can be favored if it reduces the variance in fitness and increases the geometric mean fitness. Essentially, I will try to do a better job of explaining the fitness causes and consequences of seed banks

**Line 15-17:** try not to be so terse. Explain why having some seeds not germinate immediately might reduce arithmetic mean fitness but also allow persistence across unfavorable years and thus reduce chances of extinction and variance in fitness

GS: See above.

**Line 20-21:** Are you going to be addressing predictive germination in this paper?

GS: I will reduce the emphasis on the complementary hypotheses, density-dependent models for bet hedging and predictive germination, in the Introduction. I will move discussion of these hypotheses to a later paragraph in the Introduction, and clarify their relationship to the bet hedging models I focus on in this paper. I want to place the questions asked in this paper in an appropriate context for readers who are new to the idea of bet hedging as a life history strategy for coping with environmental variability. I also expect that readers familiar with the idea may ask why the focus is purely on bet hedging, and I think bracketing other possible explanations before returning to them in the Discussion will address those concerns.

**Line 26-27:** Can you be more specific? What morphological traits and how are these correlated with bet hedging strategies, what do you mean by clines in seed behavior. Again, imagine you’re a naïve reader who isn’t familiar with the literature that you are citing. What additional information, by way of example, would help the reader understand what you are describing?

GS: Revise this paragraph to provide specific examples of morphological traits and how those are related to bet hedging. I will expand the paragraph to provide additional examples of what is known from intraspecific studies of germination delays/dormancy. A key point I want to make is that these studies demonstrate genetic or intraspecific variation in seed behavior (germination or dormancy) but have been often focused on traits rather that demography, with some exceptions (e.g. Evans 2007).

**Line 33-37:** I think the point that you are trying to make here is that the best tests of bet hedging theory have been done through inter-specific comparisons of seed dormancy behavior at a single site, but that’s not really clear. You do say the work is “in a group of species at a single site”, but that doesn’t necessarily convey the point of interspecific comparison.

I don't think it is clear from the way you describe Gremer et al's work that the point you are trying to make is that field tests of bet hedging have focused on interspecific differences in seed behavior at a single site

GS: I will emphasize that the research I cite in this section focused on inter-specific comparisons to test bet hedging. I also need to do a better job of placing this study in context of other intra-specific studies of bet hedging and distinguishing this one from that these (e.g. Philippi 1993, Clauss and Venable 2000, Evans et al. 2007).

**Line 37-39:** Would it be worth inserting something about this being an inter-population comparison (i.e., what do you mean by intraspecific variation) across an environmental gradient?

GS: Describe the intraspecific variation; in this case it’s a comparison of populations across an environmental gradient.

**Line 54:** insert "across the species' distribution in the Sierra Nevada, CA".

GS: Edit sentence to include location of populations.

**Line 51-62:** I don’t think this next paragraph is terribly clear. You tell us:

There is intraspecific variation in fitness and demography but don’t tell us what that variation is.

There is a geographic pattern in germination of first year seeds (increases from west to east).

The geographic pattern in interannual variation in growing season rain is the opposite; but in fact – CV of rainfall increases from west to east, so this is in fact the same pattern. What you mean to say is that greater interannual variation in rainfall should lead to greater interannual variation in fitness in the east relative to the west (if rainfall affects fitness), and that this in turn should favor lower germination in the east.

So, there is a crucial bit of logic that is missing from the first few sentences.

You do then tell us that fitness can vary greatly between years and, parenthetically, that this is related to variation in rainfall.

GS: Revise the paragraph to clarify the connection between environmental variation, fitness, and expected life history patterns. I will work to underscore that our questions about the variation in germination rates emerge from observations of inter-annual variation in reproductive success, and population differences in estimates of germination and dormancy. Points that need improving are the observed pattern of population growth rates, the geographic pattern of germination, and the hypothesized relationship between rainfall and fitness.

I will explicitly discuss patterns of variation in rainfall among populations, referring readers to Figure 1B-C for spring precipitation, and include similar figures for other seasons in the supplementary materials.

As I discuss the expected relationships among these variables, I will also work to make it clear that although we expect rainfall to be an important driver of fitness variation (based on observations in Clarkia, general expectations about winter annuals), it may not be the most important/only driver of fitness variation. Even if the assumption that precipitation is major determinant of reproductive success is false, we still have the fact of variation in reproductive success.

**Line 64-67:** It would really help if you addressed these predictions of bet hedging models earlier, when you first talk about bet-hedging.

GS: I discuss predictions of bet hedging models far in to the Introduction. I will integrate these predictions into my explanation of bet hedging theory (see comments on line 13-14).

Questions 4&5

GS: Questions 4 and 5 are currently dropped in without much context but we’ve discussed how to keep the focus of the introduction on bet hedging without including too many additional complications. In revising the introduction, I will keep a ‘light’ introduction to predictive germination or density dependent bet hedging, and emphasize mainly points that are connected to these 2 questions.

Methods

I was pretty confused by the methods section. I try to put myself in the place of a reader who is not at all familiar with this work and ask myself whether I could follow what is being described, and I have to conclude that I couldn’t follow. Some of this is that I really don’t know Bayesian methods to estimating parameters of interest but I think this is only part of the problem.

GS: In general, I will revise this section to better relate the goal (estimate vital rates), the data (seed bag experiment, aboveground observations), how those data are used (in analyses), and the analysis (models to estimate vital rates). I will also discuss how some observations are vital rates in and of themselves (e.g. seeds per fruit) but others require analysis to obtain the rate (e.g. germination). To make this point clearer to readers, I plan to put a figure with a life cycle diagram for *Clarkia* either in the main text or in the supplement.

Here are some points of confusion:

1. I’m not sure you do a very good job of describing what data were collected. This is especially true for the seed bag data, which admittedly are hard to describe. For example, when you first tell us about the seed bags, you say how many were put out and that they were retrieved twice (Jan and Oct) but you don’t tell us what was done when they were retrieved (what data were collected) until later in the section. It’s also not terribly obvious that a given bag is only retrieved in Jan and Oct of one year. In other words, the bags that give data on seeds in their second year aren’t the same bags that give data on seeds that are <= 1year and not the same bags that give data on seeds in their third year.

GS: I will revise this section of the manuscript. First, I will introduce the goals (parameters for model of life history) before discussing the methods and do a better job of connecting the experimental design and data to the goals. Second, I will reorganize the description of the experiment as suggested so that I give a full chronological account of one experimental round before saying that this was repeated. Third, when I discuss the data that are collected I will aim to describe how those data will be/were used in the analysis and refer readers to specific analyses.

2. In your description of the analytical methods you use y, thetas, mus, but I didn’t feel you were explicit in what these symbols refer to. So, y is for actual data, theta is latent mean of observations but also seems to be used to symbolize life history parameters (maybe?), and mu’s are never really explicitly defined.

GS: I will work to reduce the number of symbols that I use in the text, and to make it clear to readers what those symbols refer to. I will include a table that can be placed in the supplementary material that defines all the symbols in the text. I already have a table that defines priors for all parameters in the statistical models, and I will include a table that includes any symbols that refer to data or parts of the population model.

3. The same life history transition is symbolized by different symbols in different parts of the text. For example, phi first appears in the life history description and matrix to symbolize seeds per fruit, but later in the paper, phi is used to symbolize various survival transitions (either of seeds or of seedling survival to fruiting). Survival to fruiting is first symbolized in the life history/matrix section by sigma but later with phi.

GS: See above.

4. I’m also not at all sure that I understood the process of estimating vital rates for seeds. See my comment below about possible confusion with the use of the terms germination, survival, survival function, persistence, viability. But this is only part of the confusion.

You don’t really describe in the main text how you fit the survival function (is this the same as fitting a Weibull function?) The Weibull function fitting is mentioned very briefly in one of the supplemental files (I think it’s the sm-priors file). But I couldn’t understand whether this was a function fit to the intact seed count (or intact seed count+germinants) data, or how these counts were first adjusted for the viabilities measured in Oct and extrapolated back. I also wasn't sure whether you did this survival function fitting year by year (i.e., year 1 and younger seed separate from seed older than 1 year up to 2 year old seed, separate from the oldest seed), or whether data from all seed ages were used together.

And finally, I wasn't entirely clear on how this was discretized to specific time points to get the survival vital rates for particular periods (e.g., s1, s2).

GS: I will revise the section describing parameter estimation to reduce the number of terms used. In editing, I will prioritize using terminology consistently across the manuscript so that I’m not referring to seed survival in the Introduction one way and then survival in the Methods in another.

I will also cue how the data are going to be used earlier in the Methods section in order to more clearly link the data (intact seed counts, germinants) to the parts of the analysis where they are used. The ‘survival function’ is a mix of discrete and continuous components; the continuous component is the Weibull function. To avoid confusion I will be more clear that it’s a mix of discrete and continuous parts and that it represents the probability that a seed remains in the soil seed bank to a particular time. The viability data was used to adjust the estimates of survival obtained from the seed bag data alone; I will also revise my explanation of the viability data to clarify that this is how the data from lab trials are used. Finally, I fit the Weibull function to all data from the seed bags simultaneously (so, all seed ages were used together). I’m not yet sure the best way to address this because I don’t use those later ages in the analysis. But I’ll work to make this clear to readers.

**Line 77:** I don't think I've ever seen germination in October. I would say November to late February-early March

GS: Revise start of germination period from *November* to late February-early March

**Line 79**: None of our populations flowers as late as July -- it's mostly over by mid June (or earlier) because we record fruiting starting in mid June (at the earliest sites) and it takes about 3 weeks between flowering and fruit set. Some high elevation populations may flower into early July, but these aren't included in the 20 pops for which we have demographic data

GS: Change end of flowering period to *mid-June* at latest

**Line 79**: late April into early July

GS: Edit to *late* April into early July

**Line 87:** For this study, we assume that the new and old seeds differ in their survival rates in the seed bank, but do not include additional age structure and assume germination of new and old seeds is the same

GS: Edit to *…we assume that the new and old seeds differ in their survival rates in the seed bank, but do not include…*

**Line 93-95:** You use the symbol phi in multiple places in the manuscript and it means different things in different places. So, here, it’s seeds per fruit. In the section on computing vital rates, belowground rates, phi is used to refer to survival of seeds and in the section on per capita reproductive success, in the equation phi is used to refer to seedling survival to fruiting, although in this sentence and in the matrix below that is symbolized by sigma.

GS: See earlier comment on symbols.

**Line 94**: change ‘or’ to ‘and’

GS: Edit to *and* instead of *or* in this list

**Between line 95-96 [this rendered incorrectly from the PDF file]**

This is a little confusing because nothing in the equation refers to seeds that are older than 1 year.

I’m not really sure why you go through this life cycle graph and the transition matrix since you really aren’t considering data on seeds beyond their first year, and you’re not going to be using the matrix.

If you are going to include this section you need to refer the reader to the life cycle graph (you do so in the next paragraph but not here).

This needs to go earlier, when you first explain that life history transitions start in October.

GS: I cut this section from the text (see the PDF) but it still showed up in the Word document. These comments all address this specific section and I think noting that this text does/should not appear in the manuscript resolves them.

**Line 97:** I think it is important to say a little bit more about where these populations are distributed.

**Line 107**: It’s only in the first round (2005-2006) that 30 bags were put out. In the second round it was 20 bags, and in the third round it was 10 bags per site.

GS: See comment above about revisions to description of seed bag experiment. I plan to address this comment by describing the full chronological sequence of the first round before explaining to readers that the experiment was initiated in 2 subsequent years with fewer bags in those years.

**Line 108**: You haven’t told the reader what the survey plots are; people won't know what you mean by permanent survey plot, so you'll want to refer readers to some place below where you explain what these are.

GS: I could resolve this by referring readers to later in the text where I discuss the permanent plots, or by discussing the permanent plots before discussing the seed bag experiment. In either case, I will also add additional detail about the permanent plots, such as describing their number, spatial arrangement, and distribution on the landscape.

**Line 111**: I think you need to tell the reader what was done with the bags in January and October. You get to this later but it’s not going to make any sense why bags were unearthed twice each year, without being told what was done with the bags when they were unearthed.

So, I think it would be better to present the seed bag experiment more along the lines of how we described it in the 2011 paper. Start with the first round of the experiment (2005), that 30 bags were put out in October. Then what was done in the first January (scoring germinants and intact seed), including that bags were returned to the field after they were scored. What was done with the bags retrieved in the first October. Then how this was repeated for 10 more bags in the second year, and the final 10 bags in the 3rd year.

Then describe that the experiment was repeated, starting in the second year (2006) with 20 bags, which were followed for two years; and repeated again in the third year with 10 bags that were followed for one year.

I think it's important to say something about why the bags were unearthed in January and October (i.e., score germination and intact seeds in January, and estimate seed survival in October), rather than leave this to several paragraphs later. The reader is going to wonder what's going on.

Using the panels in figure 2 to illustrate the data that were collected would also be useful.

GS: My response to comments at start of Methods addresses some of these points. (1) I will reorganize my explanation of the experiment so that it follows the full chronology of the first experimental round, as suggested here. (2) I will describe the data collected when the bags are unearthed in January and October, and connect the data to the goals of the experiment. (3) I will make more use of Figure 2 in the text. I will use Figure 2A to describe the seed bag experiment and define the data that were collected. I will also use Figure 2F to more clearly show how the viability data is used to adjust estimates of seed survival.

**Line 121**: This makes it sound (or could be interpreted to sound) like intact seeds were counted in the same bag over multiple years, when this is not the case.

GS: I will revise this section to make it clear that when bags were unearthed in a given January, seeds in those bags were only counted in that January and October. I will rever to Figure 2A to emphasize this point.

**Line 130-138**: Edits to in-line description of viability trials:

GS: Edit: *At the end of each experimental year, the bags that had been scored for germination in January were brought to the lab in October and intact seeds were tested in a two-stage viability trial (Figure* [*[fig:seed-bag-experiments]*](#fig:seed-bag-experiments)*C). In the lab, we counted intact seeds and conducted germination trials and viability assays on subsets of the seeds from each bag to estimate the viability of the intact seeds.*

**Line 147:** It would be helpful to say a little bit more about these permanent plots – how they are distributed across sites (e.g., in associated documents you refer to transects, but there is nothing about transects here).

GS: See comment on Line 108.

**Line 155**: Not on all plants. We typically count fruit number per plant on up to 15 plants per plot (both permanent and additional plots). Sometimes we count fruits on more than 15 plants but generally not.

GS: Edit to clarify that fruit number was counted on up to 15 plants per plot rather than on all plants per plot.

**Line 169:** doesn't this sentence repeat what's said in the previous sentence?

GS: Yes, revise the sentence to remove the repetitive text.

**Line 176:** This section is really abstract! For someone unfamiliar with this approach it could be pretty meaningless.

Is there any way of making it more concrete/specific? For example, would it be better to show the model explicitly for one life history transition, and then say a similar approach was used for other transitions?

You also don’t mention Bayesian ever in this section.

Finally, I’m not sure what theta refers to – you call it the latent mean of observations (2nd paragraph below). So is theta jk the latent mean of some observation yjk? Or is it the latent mean of a life history transition?

GS: I will attempt to better connect this section to the data. First, I will refer readers back to the general goals that I will establish at the beginning of the Methods section (estimate parameters for life history model). Second, I will revise the text to focus on a specific dataset (likely from aboveground vital rates) and describe the relationship between the data (defined in the previous sections) and the model parameters (defined in this section). I think this is a good suggestion and similar to what is done in e.g. Elderd and Miller (2016). I will refer readers to supplementary material for similar detail on other parameters.

I was attempting to describe *theta* as the equivalent of a random intercept for population and year. I will revise my description of this point to make it more clear. I will follow the language from Metcalf et al. (2015) to be consistent with what’s used in the literature.

**Line 238-241**: I’m not sure I understand what you are getting at here.

GS: My point in this section is that winter seed survival (s1) is estimated as the product of survival and viability (i.e. survival is adjusted by viability). These are 2 datasets that are being brought together to estimate 1 parameter. We can check whether each estimate (survival alone, viability alone) is a good fit to the respective data, but we can’t say anything about whether the combined parameter s1 is a ‘good fit’ to the data because we do not have data to which we can compare this combined parameter.

**Line 242:** I think there is potential for confusion with the use of the words germination, survival and (less so) viability. You also use the word persistence (which shows up earlier in the introduction before you introduce the population matrix, and also in the header to table 2, and perhaps in one of the supplemental files).

With germination, it can be confusing because germination occurs in the bags in the field (measured in January) and then as part of the viability trials. In addition, in Fig 2E, you have several germination metrics.

GS: I see the following uses of germination: (1) germination from seeds in the field, (2) germination as part of the viability trials, (3) estimates of germination for g1, which account for viability.

I think I primarily want to use germination to refer to the parameter *g1*, as this is part of the life history model and corresponds to its use in bet hedging models. In the seed bags and viability trials, I could use the terms ‘emergence’ and ‘lab germination/emergence’.

For survival, it’s not always clear in the text whether survival simply means remaining intact (but not necessarily viable) – that seems to be the case in this next paragraph, but I wasn’t always sure.

GS: Again, I think I will primarily use survival to refer to seed survival corresponding to parameters in the life history model. So I will discuss ‘remaining intact’ when talking about the seed bags, and only introduce ‘survival’ once viability has been accounted for.

There’s also the term survival function which isn’t explained.

GS: I will switch from using the term ‘survival function’ to talking about the probability that a seed remains in the soil seed bank.

For persistence, it’s not clear whether this is the same thing as survival (remaining intact).

GS: I will edit out ‘persistence’. I agree that this is confusing when I am also using ‘survival’.

For viability, there is overall viability (the fraction of seeds tested in October that either germinated or were viable with the tetrazolium test), and there’s also viability in the tetrazolium test alone, and then there is viability interpolated to times when it wasn’t measured.

GS: I will attempt to better distinguish between viability estimated in October and interpolated/inferred viability. I will do this in part by referring readers to Figure 2D, which shows estimated and inferred viability by filled vs. empty circles. I will think about how to distinguish between overall viability and viability in tetrazolium test alone – not sure how to best do that yet.

I think it’s going to be essential that you define your terms very explicitly

**Line 242:** It seems to me that this section is really crucial, and I’m not sure it’s very accessible to readers who aren’t very sophisticated analytically.

It would be helpful, for example, if you reiterated what the data are, or preferably if you made this more explicit in the methods: For each cohort of bags (i.e bags put out in a given October, and sampled within 1, 2 or 3 years of planting), counts of geminants and intact seeds in bags in January and counts of intact seeds and estimates of their viability for the same bags the following October.

Having explained the data, making use of Fig 2 A-B, explain how the counts of intact seed in January doesn’t account for the fact that some of these are not viable, and so to estimate germination, conditional on viability, in January, you 1) assume seeds lose viability at a constant rate, starting when they were first put out, and 2) use the viability estimate of seeds assessed in the October following the scoring to germination (Fig. 2C) to extrapolate back to how many of the intact seeds were viable in January. And then show how this affects the germination probability estimates (unconditional, conditional on being intact, conditional on being viable) in Fig. 2E

**Line 242:** Do you mean germination counts? How do you use germination probabilities…..in estimates for the probability of germination?

**Line 242:** This is the first time you talk about a survival function. Is this the Weibull function you fit to the data? If it is the Weibull function, it’s buried in the supplemental file on sm-priors

GS: The ‘survival function’ is the probability that a seed remains in the soil seed bank. Because seeds can leave the seed bank through mortality or germination, loss/persistence of seeds is the combination of these processes. I represent mortality using a continuous function (the Weibull function) and germination using a discrete function (probability of germination). I will edit this section to emphasize the biological process at play rather than the conceptual details.

**Line 256:** Seems like a word is missing; do you mean “less than” or “more than”

GS: Missing word; edit to add *greater than* to the text.

**Line 260:** I’m not sure what you mean by survival function. Is this the Weibull function you’re fitting to the data?

GS: See comment above.

**Line 267:** This sentence is pretty terse and cryptic. It would be better to walk through the figures more explicitly. See my comment above.

GS: Agreed. I will also use Figure 2F to more clearly show how the viability data is used to adjust estimates of seed survival. Since I do not discuss the distinction among germination ‘types’ in the main text, I may eliminate some of the complexity in Figure 2E and only illustrate the germination that directly corresponds to what is discussed in the main text at this point.

**Line 281:** You don’t mention here the issue of sometimes recording more fruiting plants than seedlings and how you dealt with this. I can’t remember whether you deal with this is a supplemental file.

GS: I will add this section back to the manuscript main text or find an appropriate position in the supplementary material and reference it in the main text. I may have deleted it in this draft.

**Line 290:** For someone not very familiar with these Bayesian methods, this sentence isn’t going to mean anything. I believe this is the section where you want to address the issue of estimating the variance in reproductive success across years. But this only becomes (somewhat) obvious toward the end of the paragraph that follows this one.

It would help the reader if you started off this first paragraph with a topic sentence that explains the problem being confronted – how to accurately estimated among year variation in reproductive success given that years of low (or no) reproduction are years in which there are little data and the estimation procedures results in estimates that are closer to population-level means.

GS: Agreed. I will include a topic sentence as suggested. I will also include references to other work on this issue, either for estimating among year variation in vital rates (e.g. Metcalf et al. 2015) or for sampling variation in low fitness years (Evans et al. 2007)

**Line 320:** Up until this point you always use “we” but here you switch to I. I think you want to be consistent throughout.

The same thing happens in the supplemental file sm-priors, although there you start with I and then switch to we.

GS: I will revise to be consistent throughout the main text and supplements.

**Line 325**: Typically, one doesn’t start discussing/presenting results in the analysis section (same for other analyses below)

GS: I will move reference to figures and interpretation to the Results section.

**Line 352:** I don’t think I really followed what is described in this paragraph, but we can talk about it.

GS: I don’t think we discussed this in our meeting. The steps here are identical to those in Gremer and Venable (2014) and I can see how they are unclear. I will revise this section and start by discussing the goal: calculate the optimal germination fraction given the observations of reproductive success at each population. I will then reorder the details of the procedure I describe. I will start by reiterating that the ‘optimal’ G should be the one that maximizes long term geometric mean growth rate, and that the steps being taken are used to find that optimal G.

**Line 375:** It’s not obvious from this sentence that you fit a separate linear regression for each population, and that the data for each population consisted of the per capita reproductive success in each of the 15 years and the corresponding growing season precipitation for those years.

GS: Agreed. I will revise this to make what I did more clear, and to emphasize that I am fitting separate regressions to each population.

**Line 385:** Why the median? And are you saying that you used the median values of a given fitness component for each year to calculate the variance in the log of that fitness component.

GS: Yes, you’ve interpreted that correctly. For this population, I’ve calculated the median of the variance in per-capita reproductive success and partitioned that value by using the median of the fitness components. I will revise this section to make it more clear that this is what I did.

Results

**Line 410:** This statement is a little odd – that growing season precipitation doesn’t explain variation in reproductive success at any population. While it is true that none of the regressions is statistically significant (when adjusting for multiple comparisons) the proportion of the variation in RS explained by variation in precipitation is above 35% (just to use one cut-off) in 8 out of 20 populations. It think it would be better to say that “while the relationship between precip and RS is not statistically significant after adjusting for multiple comparisons, the slope….”

GS: I will edit this so that sentence states the relationship between precipitation and reproductive success is not statistically significant, but that the slope varies.

**Line 413:** The legend to Fig 7 says that it represents the variance decomposition for the geometric standard deviation of per capita reproductive success, but it seems to me that it shows the variances of the (log) of fitness components.

GS: Yes, it seems I was wrong about how I chose to label this figure. The point I was hoping to make with this is that the contribution of each fitness component to the geometric standard deviation is multiplicative, rather than additive, *and* that seedling survival to fruiting has the greatest contribution. I will consider whether I need to make both of these points and how to perhaps better use this Figure to illustrate what I want to say.

**Line 419:** I looked to see whether the pops where variance in log sigma was high were the same as the pops for which there was a strong relationship between RS and precip. There are 8 populations with high variance in log sigma (above 5). 6 out of 8 of these pops are in the top 10 pops with highest slope. Not sure what this means.

GS: I think this point is perhaps addressed by examining the correlation between seed survivorship and precipitation (see response to Figures).

Discussion

GS: In the Discussion, I will discuss the results of this study and place it in context of complementary hypotheses. If the result does not seem to support predictions of bet hedging models, what are the next steps? For example, I will describe to readers the approach one would need to take in order to test density-dependent models for bet hedging (Ellner 1985, Gremer and Venable 2014). I will also describe data that would need to be gathered in order to determine whether predictive germination (Gremer and Venable 2016) or maternal effects (Philippi 1993) are likely to explain the observed patterns. In discussing these alternatives, I will aim to more thoroughly discuss how these alternatives might be jointly expressed. In particular, I will discuss how the last 10 years have seen theory and empirical work that examines the interplay between bet hedging and predictive plasticity.

GS: If seeds germinate according to a cue, that cue should be correlated with how good of an environment the plant will experience. Winter precipitation will be a cue depending on 2 relationships: (1) the correlation between winter and spring precipitation and (2) the correlation between spring precipitation and reproductive success. These relationships seem to vary between populations; see Figure 6 for correlation between spring precipitation and reproductive success, and add figure showing correlation between winter and spring precipitation. In discussing the consequences of variation in this relationship, I will emphasize to readers that environmental variation is a mix of predictable and unpredictable variability. The relative mix of these variables should be related to whether organisms respond via bet hedging or predictive germination. However, even if the environment is predictable to the plant, it may be difficult for us as observers to identify the relevant cue(s). I will thus try to pull together the observed relationships among precipitation summaries and reproductive success to state some hypotheses about how we might expect the relative contribution of bet hedging and predictive plasticity to vary among populations.

GS: In the Discussion, I will try to discuss potential genotype by environment interactions, and how these might influence our observations and interpretation of the data. In writing about this, I will refer readers to our result that shows variability in how well precipitation predicts reproductive success. I will also cite previous work with *Clarkia* (e.g. reciprocal transplants; Eckhart et al. 2004) that has demonstrated the prevalence of GxE interactions. Furthermore, I think this is a good place to consider this study in context of others; several examine germination rates of different populations in a common environment (e.g. Philippi 1993, Clauss and Venable 2000). Some look at local rates in local conditions (e.g. Clauss and Venable 2000, Evans et al. 2007). Clauss and Venable 2000 include a discussion of how the experimental approach might confound expected patterns but I think there may be more scope to expand on this issue. To inform this part of the discussion, I will look at the relationship between winter precipitation and germination. I will plot mean g1 against mean winter rainfall to ask whether the differences by population are explained by differences in winter precipitation.

Figures

**Figure 1:**

For panel A, you need to define what the axes are.

GS: I will add longitude/latitude to the axes.

For panel D, I might also reiterate that pops are listed from furthest east (top) to furthest west (bottom).

GS: I will add text noting populations are arrayed from East (top) to West (bottom) to the figure, or note that relationship in the legend text. If I do this in the figure, I will add text above/below the populations noting East and West.

I'm not really sure about including panels E and F. As far as I could tell you don't refer to these panels in the main manuscript text, so the reader won't know why they are there and what point you're trying to make with them. If they are more relevant to material in supplemental files (e.g., the model check file), then it would probably be better to move these panels to that file.

GS: I will remove panels E and F from Figure 1. After reading comments on the paper, I think these panels do not ‘add more’ to the paper in a way that I hope they would.

**Figure 6:** I can see why you'd want to arrange pops by the magnitude of the slope but I'm wondering whether it would be better to be consistent in the ordering of populations throughout the ms (and supplemental files), i.e. arranging populations by easting from west to east.

GS: I will rearrange the populations so that they are consistently arranged by easting from west to east throughout the manuscript, including in this Figure.

**Figure 7**:

First, you describe it as a variance decomposition, but what you are really showing is the variances of the 3 fitness components.

GS: Yes, it seems I was wrong about how I chose to label this figure. The point I was hoping to make with this is that the contribution of each fitness component to the geometric standard deviation is multiplicative, rather than additive, *and* that seedling survival to fruiting has the greatest contribution. I will consider whether I need to make both of these points and how to perhaps better use this Figure to illustrate what I want to say.

Second, I think it is the variance of the log of the fitness components (according to the formula you present in the text), so the headers should be changed.

GS: I will relabel the axes.

Last, the populations are ordered alphabetically, which is the third ordering scheme used in this ms. I think it would be best to be consistent in the ordering scheme used across figures, both in the main ms and in the supplemental files, and the best ordering scheme is likely with respect to easting.

GS: Change order of populations so that they are arrayed by Easting.

Header: should this be var(log sigma), same for other headers?

GS: Additional figures that we discussed include…

1. Seed survivorship plotted against easting.
2. Germination rates plotted against easting.
3. Aboveground fitness components plotted against easting.
4. Variance in reproductive success plotted against easting.

If Figure 7 suggests that variance of seed survivorship is the biggest contributor to variance in reproductive success, what would the plot of seed survivorship versus rainfall look like?

1. I will plot seed survivorship against spring precipitation. The plot will be similar to Figure 6 in this draft of the manuscript.

Tables

**Line 174/Table 1:** The table in one of the supplementary files showing the time range of the data for aboveground vital rates shows the data range going through 2019 not 2020.

GS: Correct Table 1 in supplementary file to show that aboveground vital rates include data through 2020.

**Line 273/Table 2:** I feel like the table could be made more helpful either by expanding the legend to explain what’s being shown or adding a column with explanation. What are the theta’s in the second column (I presume they are the seed counts or seed+germinant counts) and then the third column shows how these go into estimating viable seeds.

GS: In the top half of the second column, the thetas are the parameters estimated from the seed bag experiment (so the probability of success in a binomial experiment). The third column shows how these are combined with the overall or inferred viability to get the probability that a seed survives to a certain point in time.

Why is the header for both the second and third column in the top part of the table S(xi)

GS: The header for the third column should be changed so that it reflects this difference.

For Jan1, intact, you call this theta2, but isn’t this the same as (1-gamma1) from the Jan1, total?

GS: I think that theta\_2 should be theta\_1\*(1-gamma\_1). The parameters are: theta\_1 the probability that a seed remains intact to January (before germination), gamma\_1 (the probability that a seed that is intact in January emerges as a seedling), and theta\_2 (the probability that a seed remains intact in January after germination).

Supplemental material

Data summary

In Tables 2-11, you never say what the numbers in the tables represent (except for Table 5).

GS: Add description of table contents to all Table titles, or include a legend for each table. Applies to all comments below.

**Table 1:** In the manuscript, you say that you used data through 2020, but in this table the last year in the data range is 2019

GS: Edit table to state that data includes years through 2020.

**Table 2:** What are the numbers in this table? Is it the number of seed bags that were scored?

**Table 3**: Same question. What are the numbers in this table?

**Table 4:** Again, explain what the numbers are in this table. Is it the number of plots in which there were fruiting plants?

**Table 6:** Again, explain what the numbers in the table are

Model checking

**Page 1:** Is this first section on Model checking going to be rewritten?

I'm not sure what this first section is meant to convey, as opposed to starting with the section on posterior predictive checks.

GS: I will remove the first section; it was intended to be a general overview but perhaps is not particularly effective.

**Page 5:** For almost all of the figures in this document there are no legends.

GS: I will add legends to all the figures in the document.

Only the figures near the end have legends and are numbered. And throughout the text, you don't refer to specific figures, even where figures are numbered. It would help the reader if you numbered the figures and pointed the reader to the figure that is relevant to a statement you're making in the text.

GS: I will refer to specific figures throughout the text of this supplement as appropriate.

Also, in the right hand figure panels, the acronym for the site appears (though it's sort of hidden behind the x axis values for the panels above), but the site acronyms don't show in any of the other panels.

GS: I will replot some of these to include the site acronym.

Finally, as I indicated in my comment on figures in the main text, I think it would be best to arrange populations in all of the figures from west to east -- so in these figures, the population at the top left corner would be LO and the population at the bottom right corner would be SM

GS: I will reorganize the plots so they are arrayed by easting.

**Page 12**: I really didn't understand these figures

Same for the next two figures

GS: The figure on the left hand side displays the Bayesian p-value for several test statistics. The Bayesian p-value is the “the frequency with which the test statistics calculated from the simulated datasets exceeded the test statistic calculated from the observed dataset. For example, we calculated the mean of each simulated dataset and compared this to the mean of the observed dataset. The Bayesian p-value is then the frequency with which the means of the simulated datasets exceeds the mean of the observed dataset. A similar logic applies to calculating any relevant test statistic, not just the mean. Extreme Bayesian p-values (e.g. less than 0.1 or greater than 0.9) indicate lack of model fit while values around 0.5 indicate reasonable model fit.”

So for each population, I fit a model and then generated a simulated dataset specific to that population. I then calculated the mean (for example) of all those simulated datasets and compared that mean to the mean observed in the actual data. I did this for several ‘standard’ test statistics that describe general characteristics of model fit: minimum, maximum, mean, standard deviation, Chi-squared value. Each point is the Bayesian p-value for one year\*population combination. I explain this in a legend.

The Bayesian p-value shows a lack of fit for the minimum value (values are all at extremes). The figure on the right is me simply showing that the lack of fit is an artifact of how the Bayesian p-value is calculated (frequency with which the minimums of the simulated datasets exceeds the minimum of the observed dataset).

**Page 14:** Is there some reason that you show results for 2008 and 2009?

GS: These are the model checking steps for the number of seedlings that emerge in plots. This is the dataset that contributes to estimates of s0. We only have 2 years of data in which we have estimates of fitness in plots in year t and then estimates of germination in year t+1 with intervening estimates of germination and seed survival from seed bags. That’s why there are only 2 years of results here.

What is the y axis?

GS: The y-axis is the ‘p-value’. I will revise the axis.

**Page 17:** I think the description for panel A and B should be reversed

GS: I will correct the legend description.

**Page 19:** This section doesn't seem to be about seeds per fruit and appears unfinished

GS: I will add text describing the model checks for seeds per fruit and fruits per plant.

Priors

**Page 1:** Does it make sense to start with a discussion of how the priors were set before you describe the models that were used? I think I'd prefer a description of the models first.

GS: I will better connect this section to descriptions of the models. I had intended this to be a section that stood alone in case readers wanted to understand more about the model fitting process but perhaps I need to better connect it to the main text of the manuscript.

You start using "I" but then switch to "we

GS: As in main text, I will work to be consistent in use throughout.

**Page 2:** insert "in permanent plots" after fruiting plants

GS: I will edit to include permanent plots.