Last updated: March 12, 2021

Appendix: Priors

To develop priors, I used models in published literature as well as conceptual explanations of Bayesian inference and implementation. I found it particularly helpful to understand that the amount of information encoded by prior can only be understood in the context of the likelihood. Seaman et al. 2012 and Gelman et al. 2017 provide a formulation of this point, and examples include Hobbs and Hooten 2015, p. 95-97, Wesner and Pomeranz 2020, bioRxiv, Gelman et al. 2020 - workflow, Gabry et al. 2019. p. 393-394, Northrup and Gerber 2018, Gelman et al. 2006, Gelman et al. 2008, Lemoine 2019, A1.2, STAN discourse discussion, Simpson post, NB model

I used the following general principles to identify and apply priors. First, I sought to use weakly informative priors avoided placing much probability on physically or biologically unrealistic values (hlGelman, Lemoine, Wesner and Pomeranz). Second, I used prior predictive checks to confirm that the scale of priors translated to realistic values upon parameter transformation (Hobbs and Hooten, Gabry, Wesner and Pomeranz). Third, I placed positive, unbounded priors on variance components (i.e. standard deviations), rather than priors with hard upper bounds (ref). Finally, I assessed priors by simulating prior predictive distributions Gabry et al. 2019, p. 393-394, Conn et al. 2018, p. 529-530, Hobbs and Hooten 2015, p. 85. This step helped confirm that the chosen joint likelihood (deterministic model, stochastic model) generated data within the observed range. This is similar in logic to the approach taken by Evans et al. 2010, Methods: Estimating vital rates from demographic data: Priors except in that case the authors compared their observed means to those generated by their priors.

The 'default' flat priors recommended for linear models are not always noninformative for models with different likelihood functions. Because none of my likelihood functions were normal (binomial, negative binomial), I identified priors that were relatively noninformative in the context of the joint likelihood for each model. References for these priors include the following: Hindle et al. 2019, Table S1, Rosenbaum et al. 2019, Table 1, Hobbs et al. 2015, Table 3, Hindle et al. 2018, Smits 2015. I also followed the guidance in Lemoine 2019 to use positive, unbounded priors on variances, and to use Cauchy priors for the random-intercepts.

I parameterized my models with hierarchical centering Ogle and Barber 2020, Evans et al. 2010, Table 3 and Appendix A. This is equivalent to random effects formulation but structures the prior in a similar way to the data (populations, years within populations).

Binomial likelihood

We modeled counts of germinants in the seed bag trials and counts of fruiting plants as a binomial likelihood, with a logit-link and hierarchical normal parameterization.

The parameters for population means ($\mu^{\text{germ}}_{0,ik}$; ...) were given $N(0,1)$ priors [ref: β^g_0, β^s_0, Evans et al. 2010, Table 3]. The parameters for the standard deviation of population and population-and-year levels of the hierarchy (...) were given $N(0,1)^+$ priors [ref: $\sigma_{...}$, Rosenbaum et al. 2019, Table 1]. I chose to apply half-Normal $N(\dots,\dots)^+$ priors to the standard deviations to not restrict the support of the prior to a particular range. Finally, I used prior predictive checks to assess whether the priors were relatively uninformative on the probability scale before proceeding to use them for inference.

Binomial likelihood, deterministic survival process

We modeled counts of seeds in the seed bag trials with a deterministic Weibull survival function. We used a binomial likelihood and a logit-link. We used a hierarchical normal parameterization for the inverse-scale parameter, and placed a prior directly on the population-level shape parameter.

The shape parameter of the Weibull survival function (α_j) was given a $gamma(2,2)$ prior. I made this choice because the distribution has positive support and approaches 0 smoothly. The parameters for the population and population-and-year levels of the inverse-scale parameter were given $N(0,1)^+$ priors [ref: $\sigma_{...}$, Rosenbaum et al. 2019, Table 1]. I chose to apply half-Normal $N(\dots,\dots)^+$ priors to the standard deviations to not restrict the support of the prior to a particular range. Finally, I used prior predictive checks to assess whether the priors were relatively uninformative on the probability scale before proceeding to use them for inference.

Poisson likelihood

We modeled counts of fruits per plant and seeds per fruit as a Poisson likelihood, with a log-link and hierarchical lognormal parameterization.

The parameters for population means ($\nu_{\text{TF}}, \nu_{\text{UF}}, \nu_{\text{DF}}, \nu_{\text{US}}, \nu_{\text{DS}}$) were given $gamma(1,1)$ distributions. We placed $N(0,1)^+$ distributions on the standard deviation parameters (e.g. $\sigma^2_{\text{TF}}, (\sigma^{\text{pop}}_{\text{TF}})^2$).

Identifying reasonable priors for models with a log-link was a bit trickier. The log-link exponentiates parameters, which means that changes to the parameters on the latent scale have a multiplicative effect. We considered two issues: computational stability and biological realism. Overly broad prior distributions for the latent

variable in models with a log-link can lead to numerical instability in the MCMC sampler (Evans et al. 2010). The largest value that R can store is $2e + 308$ (see 'help("double")') and $\log(2e + 307) \approx 707$. A 'vague' prior of $N(0, 1000)$ will draw values on the latent scale that R is unable to store on the transformed scale. Such broad priors can also be biologically unrealistic, clashing with reasonable intuition about a study system. Wesner and Pomeranz, 2020 illustrate this effect in a model for counts of spiders in 2.32 m^2. They examine a model with a Poisson likelihood, and log-link linear model. A $N(0, 10)$ prior would typically be considered a relatively weak prior; with a Poisson likelihood this prior places a not-insignificant probability on observing $> 100,000$ spiders and a small probability on observing > 100 million spiders.

To set our priors, we followed the general approach recommended by Polson 2012, Gelman et al. 2017, Gabry et al. 2019, Wesner and Pomeranz 2020 and set priors that were computationally stable, exhibited desired properties, and generated predictive prior distributions that reflected biologically realistic counts of fruits and seeds. The parameters for population means $(\nu_{\text{TF}}, \nu_{\text{UF}}, \nu_{\text{DF}}, \nu_{\text{US}}, \nu_{\text{DS}})$ were given $gamma(1, 1)$ distributions. Any probability mass on values less than 0 on the latent scale is mapped to $(0, 1)$ once exponentiated; a $gamma(1, 1)$ retains some probability mass on small values but shifts the overall distribution by a factor of $\exp(1)$ on the transformed scale. For the standard deviation parameters, we used unbounded, truncated normal distributions so as to not set an upper limit on the parameter. We also considered a half-Cauchy or Student-t, but the tails of these distributions are thicker and gave higher latent means that were unrealistic.

Across all datasets on seed and fruit counts, the maximum count was 282 fruits on an undamaged plant and the maximum 99.9th percentile was ~ 112.2 in the number of undamaged fruits per plant. For the combination of priors we used, the 99.9th percentile of the prior predictive was ≈ 2700 (UPDATE). The priors for this model place a small probability on observing a plant that has over 10 times as many fruits as the largest plant observed in the dataset.