

## Priors

To develop priors, I used models in published literature as well as a conceptual understanding of Bayesian inference and implementation.

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

Priors were assessed 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.

## General principles

- apply weakly informative priors (Gelman, Lemoine)
- place positive, unbounded priors on the variance components/standard deviations
- check scale of untransformed and transformed parameters (Hobbs and Hooten, Gabry, Wesner)

- model with hierarchical centering (Evans 2010, Ogle 2020)
- use prior predictive checks (Conn)

## 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 ( $\mu_{0,ik}^{\text{germ}}; \dots$ ) were given  $N(0, 1)$  priors [ref:  $\beta_0^g, \beta_0^s$ , Evans et al. 2010, Table 3]. The parameters for the standard deviation of population and population-and-year levels of the hierarchy ( $\dots$ ) were given  $N(0, 1)^+$  priors [ref:  $\sigma_{\dots}$ , Rosenbaum et al. 2019, Table 1].

Initially, I placed a  $N(0, 1000)$  prior on the population means and a  $\text{Uniform}(0, 1.5)$  distribution as a prior on the parameters for standard deviation for population and population-and-year levels of the hierarchy. The normal distribution on the population mean was motivated by my understanding that it was a noninformative, vague prior. The distribution on the standard deviations were based on literature (e.g. AHM; Okamoto et al. 2016, Table B1, among-site variance; Evans et al. 2010, Table 3; Eldred and Miller 2015). I chose to apply half-Normal priors to the standard deviations to not restrict the prior to a particular range. I used prior predictive checks to assess whether the prior covered similar parameter space as the uniform prior. Finally, I used prior predictive checks to assess whether the priors were reasonable and felt comfortable applying priors not copied from the literature or defaults.

## Binomial likelihood, deterministic survival function

Gamma(2) prior or a log-normal prior log-normal(0,1) or log-normal(0,SD=.5) The benefit of the Gamma(2,2) is that it approaches 0 smoothly, while the log-normal places very little mass on 0. The geometry of the Gamma(2,2) is a bit more unwieldy but both give similar upper and lower bounds so I think should work.

Question remains about prior on SD parameter (dnorm(0,1) or dnorm(0,2) or dnorm(0,sqrt(10))

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 ( $\mu_{0,ik}^{\text{germ}}; \dots$ ) were given  $N(0, 1)$  priors [ref:  $\beta_0^g, \beta_0^s$ , Evans et al. 2010, Table 3]. The parameters for the standard deviation of population and population-and-year levels of the hierarchy ( $\dots$ ) were given  $N(0, 1)^+$  priors [ref:  $\sigma_{\dots}$ , Rosenbaum et al. 2019, Table 1].

Initially, I placed a  $N(0, 1000)$  prior on the population means and a  $\text{Uniform}(0, 1.5)$  distribution as a prior on the parameters for standard deviation for population and population-and-year levels of the hierarchy. The normal distribution on the population mean was motivated by my understanding that it was a noninformative, vague prior. The distribution on the standard deviations were based on literature (e.g. AHM; Okamoto et al. 2016, Table B1, among-site variance; Evans et al. 2010, Table 3; Eldred and Miller 2015). I chose to apply half-Normal priors to the standard deviations to not restrict the prior to a particular range. I used prior predictive checks to assess whether the prior covered similar parameter space as the uniform prior. Finally, I used prior predictive checks to assess whether the priors were reasonable and felt comfortable applying priors not copied from the literature or defaults.

For the survival models, I used a normal for the rate parameter (should this be a truncated normal?). [ref:  $\phi$ ..., Shriver et al. preprint, Table S1]. I used truncated, normally distributed priors for the parameters for standard deviation for population and population-and-year levels of the hierarchy. Smits 2015 places a half-Cauchy prior on the variance term of the random intercept but when I evaluated this prior in the context of the binomial likelihood I found it to have too thick of a tail.

We construct a model to relate all the sources of data (cf. Metcalf et al. 2009, 2.8: Prior specification)

SIMULATE PRIORS FOR EXPONENTIAL DECAY MODEL; see Makinen and Vanhatalo 2017 for the inverse scale parameter

## Poisson likelihood

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

The parameters for population means (...) were given  $N(1, 1)$  priors. The parameters for the standard deviation of population and population-and-year levels of the hierarchy (...) were given  $N(0, 1)^+$  priors.

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<sup>2</sup>. 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. For the mean parameters, we used a normal distribution with a mean of 1 and standard deviation of 1. Any probability mass on values less than 0 on the latent scale is mapped to  $(0, 1)$  once exponentiated; a  $N(1, 1)$  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  $\sim 112.2$  in the number of undamaged fruits per plant. For the combination of priors we used, the 99.9th percentile of the prior predictive was  $\approx 2700$ . 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.

We note that we found several examples of models with a Poisson likelihood that used  $N(0, 1)$  priors ([Smits 2015](#), [de Vemeruiel et al. 2020](#)). However, these models were fit using STAN, which has the option of parameterizing the Poisson on the log scale ('poisson\_log'). This parameterization conveniently allows users to set a  $N(0, 1)$  prior on the log of the mean of the latent variable, and a  $N(0, 1)$  prior on the log of the SD of the latent variable. To our knowledge, this parameterization is not available for JAGS. But use of a default prior doesn't obviate the need to assess whether that prior is plausible, and the approach we take to set priors is consistent with recommendations on this part of the model fitting workflow [Gelman workflow](#).

Initially, I placed a  $N(0, 1000)$  prior on the population means and a  $\text{Uniform}(0, 1.5)$  distribution as a prior on the parameters for standard deviation for population and population-and-year levels of the hierarchy. The normal distribution on the population mean was motivated by my understanding that it was a noninformative, vague prior. The distribution on the standard deviations were based on literature (e.g.

Evans et al. 2010, Table 3; Eldred and Miller 2015).

## Model comparison

I used leave-one-out cross validation to compare models. A lower elppd indicates better predictive accuracy. The standard error of the difference in elpd between the two models can be used to assess whether one model has better out of sample performance than another [Evans et al. 2017] Difference is 2.1, SE 3.3 so predictive performance is indistinguishable.

See Appendix S1 in Feist et al. 2017 for implementation of K-fold cross validation

## Negative binomial likelihood

CHECK IF THIS WORKS FOR NEGATIVE BINOMIAL; FIT DISPERSION PARAMETER

The models for fruits per plant and seeds per fruit use a negative binomial likelihood. For these models, I initially used a  $N(0, 1000)$  distribution as a prior for the population mean, and a  $\text{Uniform}(0, 1.5)$  distribution as a prior on the parameters for standard deviation for population and population-and-year levels of the hierarchy. See above paragraph for explanation. An alternative prior for the population mean is  $N(0, 5)$  [ref:  $\beta_3$ , Shriver et al. 2019, Table S1], which is on the weakly-informative side. An alternative for the standard deviation is a half-Cauchy (0,1) [ref:  $\tau$ , Shriver et al. 2019, Table S1]

The negative binomial likelihood also takes a dispersion parameter,  $\kappa$ , that I gave a  $\text{gamma}(0.001, 0.001)$  prior distribution [ref: Hobbs and Hooten 2015, p. 253]. An alternative prior for the dispersion parameter would be a half-Cauchy(0,5) [ref:  $\kappa$ , Shriver et al. 2019, Table S1]. Both these options are positive, unbounded.

Schultz et al. (2017) use a non-centered parameterization but use  $N(0, 10)$  priors on means, half-Cauchy(0,2.5) on the standard deviations for random intercepts, and a  $\text{IG}(0.001, 0.001)$  on the overall standard deviation. [See Appendix S1 for the growth model]

Another Poisson model:  $N(0, 3.16)$  for intercept of log of mean fecundity, which I assume is related to a  $N(0, \text{var}=10)$ .  $U(0, 1)$  for the SD of random effect on log fecundity. Drummond et al. 2018, Table S1

Makinen and Vanhatalo 2017 also include details on priors for NB model, as well as for the inverse scale parameter

Bjork et al. 2018, 2.1.1 describe the use of half-Cauchy(0,2.5) for location parameters and half-Cauchy(0,1) for variance parameters.

Next steps: identify reference for the parameter in the beta distribution. Identify reference for parameters in the rate parameter for the exponential survivorship function. Identify references for the parameters in the hierarchy of the joint likelihoods with negative binomial likelihood.

CONSIDER: change population means to Cauchy because I \*do\* have small sample sizes sometimes;

Saturday:

1. write NB model with the original and revised priors, visualize prior predictive distributions 2. Write binomial model with original and revised priors, visualize prior predictive distributions 3. write binomial model with exponential decay process, visualize prior predictive distributions

Notation in all parameters is for site  $i$ , year/round  $j$ , age  $k$ . Superscript plus indicates a distribution truncated to be positive

TABLE 1: Description of parameters for prior distributions at the highest level (hyperparameters).

Parameter	Description	Distribution	Type
SEED BAG BURIAL EXPERIMENTS			
$\mu_{0,ik}^{\text{germ}}$	Population mean germination	$N(0, 1)$	Weakly informative
$\sigma_{0,ik}^{\text{germ}}$	Population S.D. of germination	$N(0, 1)^+$	Weakly informative
$\sigma_{0,ijk}^{\text{germ}}$	Population and year S.D. of germination	$N(0, 1)^+$	Weakly informative
$\mu_{0,i}^{\text{decay}}$	Population mean <u>rate parameter</u>	$N(0, 1)$	Weakly informative
$\sigma_{0,i}^{\text{decay}}$	Population S.D. of <u>rate parameter</u>	$N(0, 4)^+$	Weakly informative
$\sigma_{0,ij}^{\text{decay}}$	Population and year S.D. of <u>rate parameter</u>	$N(0, 4)^+$	Weakly informative
$\beta_{ij}^{\text{decay}}$	Population and year shape parameter of beta	$N(0, 1000)^+$	Vague
SEEDLING SURVIVAL TO FRUITING			
$\mu_{0,i}^{\text{surv}}$	Population mean seedling survival	$N(0, 1)$	Weakly informative
$\sigma_{0,ij}^{\text{surv}}$	Population S.D. of seedling survival	$N(0, 1)^+$	Weakly informative
$\sigma_{0,ij}^{\text{surv}}$	Population and year S.D. of seedling survival	$N(0, 1)^+$	Weakly informative
NEGATIVE BINOMIAL LIKELIHOOD: OPTION 1			
$\mu_0^{\text{seeds}}$	Population mean fruits per plant	normal(0, 1000)	
$\sigma_0^{\text{seeds}}$	Population S.D. of fruits per plant	uniform(0, 1.5)	
$\kappa$	Dispersion parameter	gamma(0.001, 0.001)	Hobbs and Hooten
NEGATIVE BINOMIAL LIKELIHOOD: OPTION 2			
$\mu_0^{\text{seeds}}$	Population mean	$N(0, 5)$	Shriver 2019
$\sigma_0^{\text{seeds}}$	Population S.D.	half-Cauchy(0,1)	Shriver 2019
$\kappa$	Dispersion parameter	half-Cauchy(0,5)	Shriver 2019
POISSON LIKELIHOOD: OPTION 3			
$\mu_0^{\text{seeds}}$	Population mean	$N(1, 1)$	see text
$\sigma_0^{\text{seeds}}$	Population S.D.	$N(0, 1)$	see text