

# My, how you've grown: a practical guide to modeling size transitions for Integral Projection Model (IPM) applications

Tom E.X. Miller<sup>\*a</sup> and Stephen P. Ellner<sup>b</sup>

<sup>a</sup>Department of BioSciences, Rice University, Houston, TX

<sup>b</sup>Department of Ecology and Evolutionary Biology, Cornell University,  
Ithaca, New York

**Running header:** Better growth modeling for IPMs

---

\*Corresponding author. Department of BioSciences, Rice University, Houston, TX 77005-1827. Email:  
tom.miller@rice.edu Phone: 713-348-4218

<sup>1</sup> **Abstract**

<sup>2</sup>    1.

<sup>3</sup>    2.

<sup>4</sup>    3.

<sup>5</sup>    4.

<sup>6</sup> **Keywords**

## 7 Introduction

8 Structured demographic models – matrix and integral projection models (MPMs and  
9 IPMs) – are powerful tools for data-driven modeling of population dynamics and viabil-  
10 ity that are widely used in basic and applied settings. In contrast to MPMs for popula-  
11 tions with discrete structure (life stage, age class, etc.), IPMs (Easterling et al., 2000) read-  
12 ily accommodate populations structured by continuous state variables, most commonly  
13 size. A related innovation of the IPM framework is its emphasis on regression-based  
14 modeling for parameter estimation, which carries important advantages for making the  
15 most of hard-won data (Ellner et al., 2022).

16 A standard workflow allows ecologists to assemble an IPM from data using famili-  
17 iar statistical tools to describe growth, survival, reproduction, and other demographic  
18 transitions as functions of size (Coulson, 2012; Ellner et al., 2016). The relative ease of  
19 the regression-based approach, accommodating multiple covariates (e.g., environmental  
20 factors, experimental treatments) and complex variance structures (e.g., random effects,  
21 correlated errors), has facilitated a growing body of IPM literature that examines how  
22 biotic or abiotic factors affect population dynamics (e.g., Louthan et al., 2022; Ozgul  
23 et al., 2010; Schultz et al., 2017) and explores the consequences of demographic hetero-  
24 geneity associated with spatial, temporal, and individual variation (e.g., Compagnoni  
25 et al., 2016; Crone, 2016; Plard et al., 2018). The vital rate regressions (or “sub-models”)  
26 are the bridge between the individual-level data and the population-level model and its  
27 predictions; it is important to get them right.

28 Compared to other vital rates, growth is special. The regression sub-models for  
29 survival and reproduction provide the expected values of those rates as functions of  
30 size (we use “size” as the name for whatever continuous variable defines the population  
31 structure, which could instead be immune competence, mother’s weight, etc.). However,  
32 for modeling growth, the full probability distribution of subsequent size, conditioned on  
33 initial size, must be defined. This distribution defines the growth ‘kernel’  $G(z', z)$  that  
34 gives the probability density of any future size  $z'$  at time  $t + 1$  conditional on current size  
35  $z$  at time  $t$ . Whenever survival and reproduction are size-dependent, the entire distribu-  
36 tion of size transitions can strongly influence IPM predictions because this distribution  
37 governs how frequently size changes are much greater or much lower than average.

38 The original template for modeling size transitions in IPMs was provided by East-  
39 erling et al. 2000. They first tried simple linear regression, assuming normally dis-  
40 tributed size changes with constant variance. Because the residuals from this regression  
41 exhibited non-constant variance, they used a two-step approach that estimated the size-

42 dependence in the growth variance (better options soon became available, such as the  
43 `lme` function in R). However, even after accounting for non-constant variance, growth  
44 data may still deviate from the assumption that size transitions are normally distributed.  
45 Size transitions are often skewed such that large decreases are more common than large  
46 increases (Peterson et al., 2019; Salguero-Gómez and Casper, 2010), or vice versa (Stub-  
47 berud et al., 2019). Size transitions may also exhibit excess kurtosis ('fat tails'), where  
48 extreme growth or shrinkage is more common than predicted by the tails of the normal  
49 distribution (Hérault et al., 2011).

50 The observation that the normal distribution may poorly describe size transitions  
51 in real organisms has been made before, and several studies have emphasized that al-  
52 ternative distributions should be explored (Easterling et al., 2000; Peterson et al., 2019;  
53 Rees et al., 2014; Williams et al., 2012). Yet, default use of Gaussian growth distribu-  
54 tions (often with non-constant variance) remains the standard practice. An ISI Web of  
55 Knowledge search on the terms 'integral projection model ecology' (DATE) returned #  
56 IPM studies published in the past decade (2010–2020), # of which assumed a Gaussian  
57 growth kernel.<sup>1</sup> The general state-of-the-art in the literature appears to remain where it  
58 was 20 or so years ago, using the default model without pausing to examine critically  
59 whether or not it actually provides a good description of the data. We are guilty of this,  
60 ourselves.

61 The persistence of Gaussian growth modeling is understandable. There is a long  
62 tradition of statistical modeling built on the assumption of normally distributed residu-  
63 als with constant variance. Popular software packages such as `lme4` (Bates et al., 2007)  
64 and `MCMCglmm` (Hadfield et al., 2010) make it easy to fit growth models with po-  
65 tentially complex fixed- and random-effect structures, but the possible distributions of  
66 continuous responses are limited, and default to Gaussian. Abandoning these conve-  
67 nient tools for the sake of more flexible growth modeling means, it may seem, sacrificing  
68 the flexibility to rigorously model diverse and potentially complex sources of variation  
69 in growth, some of which may be the motivation driving the study in the first place.

70 The question we address here is: how can ecologists escape the apparent trade-off  
71 between realistically capturing the variance, skew, and kurtosis of size transition data  
72 on the one hand, and flexibly including the multiple covariates and random effects that  
73 often have substantial impacts on demographic rates. In this article, we offer an answer.

74 Our goal here is to present and illustrate a general 'recipe' that moves growth mod-  
75 eling past the standards set over 20 years ago. Like any recipe, users may need to  
76 make substitutions or add ingredients to suit their situation. Our approach emphasizes

---

<sup>1</sup>I still intend to do this! But it's a rabbit hole I have not gone down yet.

77 graphical diagnostics for developing and evaluating growth models, rather than a pro-  
78 cess centered on statistical model selection. Through a set of empirical case studies we  
79 demonstrate how a simple workflow, using tools that were nonexistent or not readily  
80 available when IPMs first came into use, makes it straightforward and relatively easy to  
81 identify when the default model is a poor fit to the data, and to then choose and fit a  
82 substantially better growth model that is no harder to use in practice. We illustrate our  
83 approach by revisiting four of our own, mostly published IPM analyses that assumed  
84 Gaussian growth.<sup>2</sup> In each case, the Gaussian assumption does not stand up to close  
85 scrutiny. We illustrate how we could have done better, and the consequences of “doing  
86 better” for our ecological inferences. All of our analyses may be reproduced from code  
87 and data that are publicly available (see Data accessibility statement).

## 88 A general workflow for better growth modeling

89 The modeling workflow that we suggest runs as follows (Fig. 1):

- 90 1. *Fit a “pilot” model or models assuming a Gaussian distribution but allowing for non-*  
91 *constant variance.*

92 This step is familiar to most IPM users, as it is the start and end of the traditional  
93 workflow. A well-fitted Gaussian model accurately describes the mean and variance  
94 of future size conditional on current size and possibly on other measured covariates  
95 or random effects. This step may include model selection to identify which treat-  
96 ment effects or environmental drivers affect the mean and/or variance of future size.  
97 Non-constant variance is often fitted in a two-stage process, first fitting mean growth  
98 assuming constant variance, then doing a regression relating the squared residuals  
99 from the initial fit to the fitted mean. It is sometimes better to fit size-dependence  
100 in the mean and variance simultaneously, as can be done with the R packages **mgcv**  
101 and **nmle**, because incorrectly assuming constant variance can affect the outcome of  
102 model selection for the mean. One-step fitting is straightforward for simple models  
103 in which initial size is the only factor that can influence growth variance. However,  
104 the two-step process fitting residuals to the fitted value (expected future size) may  
105 be convenient when there are multiple fixed and random effects, all of which may  
106 contribute to non-constant variance, since the expected value implicitly accounts for  
107 all of them. We illustrate both one-step and two-step approaches in the examples  
108 below.

---

<sup>2</sup>Need to commit to case study choices - Steve wanted to include corals for contrast with Peterson et al.

109 Allowing non-constant variance means that it is not necessary to transform the  
110 data in a way that stabilizes the growth variance. Transformation remains an option  
111 when it does not create new problems (see Discussion), and it may have advantages  
112 besides variance stabilization. In particular log-transformation is often appropriate  
113 for size data (Ellner et al., 2016), and it helps avoid eviction at small sizes.

- 114 2. *Use statistical and graphical diagnostics to identify if and how the standardized residuals*  
115 *deviate from Gaussian, and to identify a more appropriate distribution.*

116 If the Gaussian pilot model is valid, the set of standardized residuals (standardized  
117 by the standard deviation) should be Gaussian with mean zero and unit variance,  
118 with no skew or excess kurtosis. This criterion provides a straightforward test for  
119 whether to accept a Gaussian growth model or explore alternatives. If the standard-  
120 ized residuals are satisfactorily Gaussian, skip to the final step of the workflow.

121 There are many ways that growth data may deviate from Gaussian, and the  
122 nature of those deviations can guide the search for a better distribution. Frequentist  
123 tests such as the D'Agostino test of skewness (D'Agostino, 1970) and the Anscombe-  
124 Glynn test of kurtosis (Anscombe and Glynn, 1983) could be used to diagnose  
125 whether the aggregate distribution of standardized residuals deviates from normal-  
126 ity (R package **moments** (Komsta and Novomestky, 2015)). However, the aggregate  
127 distribution of standardized residuals may be misleading if properties such as skew  
128 and kurtosis vary with size. For example, a change in the direction of skewness from  
129 small to large sizes would require a distribution flexible enough to accommodate  
130 both positive and negative skew, such as the skewed normal or Johnson  $S_U$  distri-  
131 butions. Alternatively, growth data may lack skew but may exhibit leptokurtosis (in  
132 which case the  $t$  distribution may be a good choice) or may shift from platykurtos-  
133 sis to leptokurtosis depending on initial size (in which case the power exponential  
134 distribution may be a good choice). It is therefore essential to visualize trends in dis-  
135 tribution properties with respect to size, either initial size (for simple models with  
136 only size-dependence) or expected future size (for models with multiple fixed ef-  
137 fects). In the case studies below, we rely on quantile regression of the standardized  
138 residuals to visualize skew and kurtosis as continuous functions of size or expected  
139 value. Fig. 1 includes guidance on how the skew and kurtosis properties of the stan-  
140 dardized residuals suggest options for an appropriate growth distribution. In our  
141 case studies we take advantage of the many distributions provided in the **gamlss** R  
142 package (Stasinopoulos et al., 2007), but any other distributions with the necessary  
143 properties can be used.

144 3. *Refit the growth model using the chosen distribution.*

145 In models with multiple covariates and/or random effects, each potentially affecting  
146 several distribution parameters (location, scale, skew, kurtosis) in different ways,  
147 “refit the model” could entail a massive model selection process to identify the  
148 “right” or “best” non-Gaussian model. And with so many options, model uncer-  
149 tainty may be overwhelming and over-fitting becomes a significant risk even if pre-  
150 cautions against it are taken. We therefore argue for adopting the more modest  
151 goal of remedying the apparent defects in the Gaussian model. Conveniently, as  
152 we demonstrate below, the functional forms for the mean and standard deviation  
153 (or location and scale parameters) could be carried over from the pilot Gaussian  
154 model into a non-Gaussian distribution, leaving skew and kurtosis as the targets for  
155 improvement. This step exploits the fact that parameter estimation from a Gaus-  
156 sian model is generally robust to deviations from normality (Schielzeth et al., 2020),  
157 meaning that the mean of the Gaussian model is probably a good proxy for the mean  
158 of the non-Gaussian model (and in case it is not, the next step in the workflow would  
159 catch that). The functional forms for skew and kurtosis of the non-Gaussian model  
160 can be guided by the qualitative features of the graphical diagnostics (e.g., skewness  
161 switches from positive to negative with size).

162 4. *Test the final model through graphical diagnostics comparing simulated and real growth data.*

163 A good model will generate simulated data that look like the real data. Again, it is  
164 important to inspect the properties of simulated data conditional on present size or  
165 expected future size, rather than examining the entire distribution. We provide ex-  
166 amples below of informative comparisons between simulated and real data, based  
167 mainly on quantiles. If the simulated data do not correspond well with real data,  
168 alternative (possibly more flexible) growth distributions should be explored, or more  
169 complex functions relating distribution parameters to current size and other covari-  
170 ates. However, we again caution against launching a full-blown model selection  
171 exercise. Instead, possible alternative models could be chosen primarily to remedy  
172 observable discrepancies between the real and simulated size transition data, and at  
173 most slightly modified based on final diagnostic and statistical tests.

## 174 How should skewness and kurtosis be measured?

175 “Improvement” of a Gaussian model will always involve scrutiny of skewness and kur-  
176 tosis, so measurement of these properties warrants some attention. The standard mea-

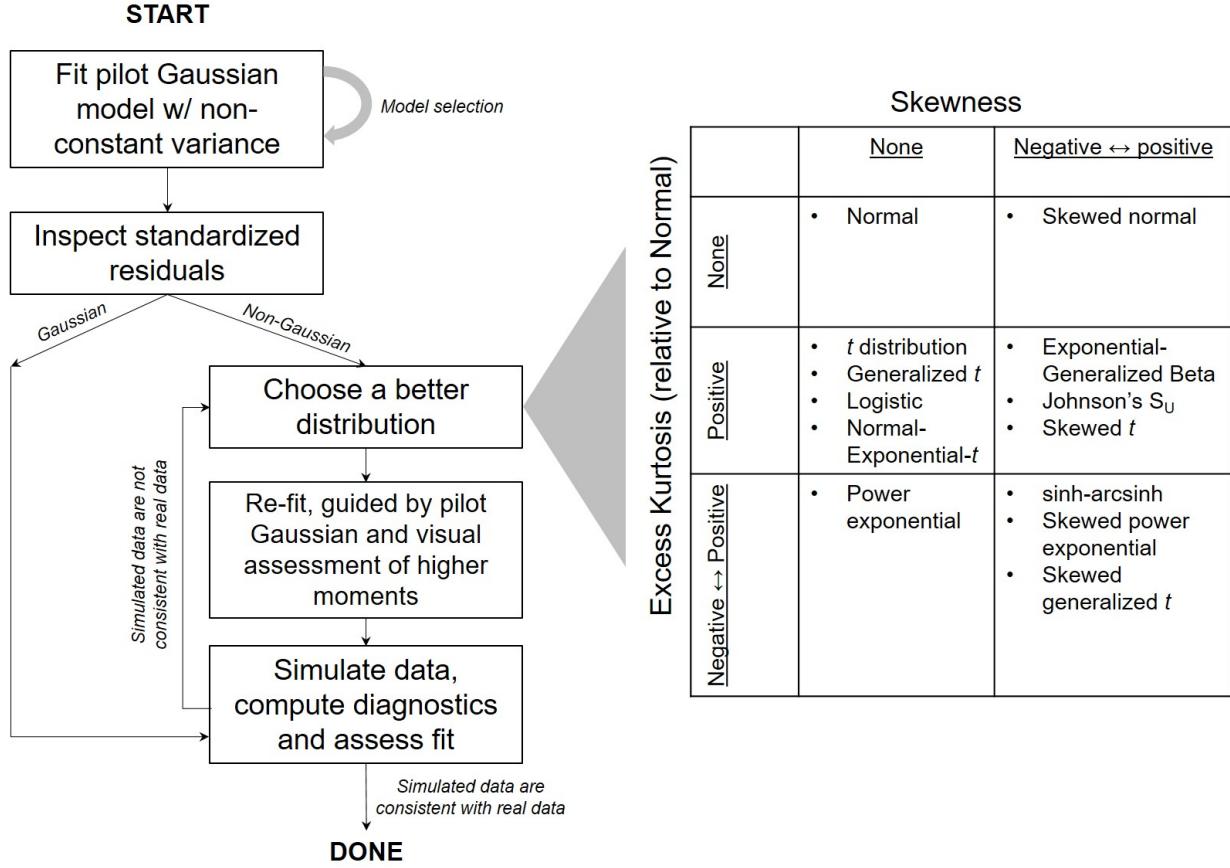


Figure 1: General workflow of recommendations for IPM growth modeling (left) and guide to common non-Gaussian distributions of size  $x$  for  $x \in \mathbb{R}$  that can accommodate different combinations of skewness and kurtosis (right). All of these distributions (often including multiple versions or parameterizations of each) are available in the package **gamlss.dist**, except for the skewed generalized *t*, which is available in the package **sgt** (Davis, 2015).

<sup>177</sup> sures of skewness and kurtosis (tail thickness) are based on the third and fourth central  
<sup>178</sup> moments, respectively, of the distribution:

$$\text{Skewness} = \frac{m_3}{\sigma^3}, \quad \text{Excess kurtosis} = \frac{m_4}{\sigma^4} - 3 \quad (1)$$

<sup>180</sup> where  $m_k = \mathbb{E}(X - \bar{X})^k$  is the  $k^{th}$  central moment of a random quantity  $X$  and  $\sigma^2$  is the  
<sup>181</sup> variance (second central moment). A Gaussian distribution has zero skewness and zero  
<sup>182</sup> excess kurtosis.

<sup>183</sup> The standard measures are easy to calculate but their use for choosing and eval-  
<sup>184</sup> uating growth models is hindered by their poor sampling properties. Because empirical  
<sup>185</sup> estimates involve high powers of data values, it only takes a few outliers to produce



Figure 2: Histograms of skewness and kurtosis estimates using moment-based definitions, compared with the nonparametric measures. Histograms are based on 5000 replicate draws of a sample of 200 independent values from a  $t$  distribution with 8 degrees of freedom. Dotted red vertical lines mark the minimum and maximum of sample estimates, and dashed blue lines show the 5th and 95th percentiles. The true value is indicated by a black dot on the  $x$ -axis. Figure drawn by script `NPmoments.R`

186 a very inaccurate estimate. Figure 2 shows a simulated example, where the underlying  
 187 “data” are a sample of size 200 from a  $t$  distribution with 8 degrees of freedom; the true  
 188 skew is 0, and the true excess kurtosis is 1.5. The distance between the largest and small-  
 189 est estimates (indicated by the dotted red vertical lines), relative to the distance between  
 190 the 5th and 95th percentiles, shows the broad extent of extreme values that can occur  
 191 even with a good size sample, especially for kurtosis.

192 We therefore use “nonparametric” (NP) measures of skew and kurtosis that are  
 193 based on quantiles and thus less sensitive to a few extreme data values. Let  $q_\alpha$  denote  
 194 the  $\alpha$  quantile of a distribution or sample (e.g.,  $q_{0.05}$  is the 5th percentile). For any  
 195  $0 < \alpha < 0.5$ , a quantile-based measure of skewness is given by (McGillivray, 1986)

$$196 \text{NP Skewness} = \frac{q_\alpha + q_{1-\alpha} - 2q_{0.5}}{q_{1-\alpha} - q_\alpha}. \quad (2)$$

197 NP Skewness is a measure of asymmetry between the tails of the distribution above and  
 198 below the median. The size of the upper tail can be measured (for any  $0 < \alpha < 0.5$ ) by  
 199  $\tau_U = q_{1-\alpha} - q_{0.5}$ ; for  $\alpha = 0.05$  this is the difference between the 95th percentile and the  
 200 median. The lower tail size is  $\tau_L = q_{0.5} - q_\alpha$ . The definition above is equivalent to

$$201 \quad \text{NP Skewness} = \frac{\tau_U - \tau_L}{(\tau_U + \tau_L)}. \quad (3)$$

202 So an NP Skewness of  $\pm 0.2$  says that the difference in tail sizes is 20% of their total. The  
 203 range of possible values is -1 to 1. Both  $\alpha = 0.25$  (sometimes called “Kelly’s skewness”)  
 204 and  $\alpha = 0.1$  (“Bowley’s skewness”) are common choices. We used  $\alpha = 0.1$ , unless  
 205 otherwise stated.

206 An analogous quantile-based measure of kurtosis (Jones et al., 2011) is

$$207 \quad \text{NP Kurtosis} = \frac{q_{1-\alpha} - q_\alpha}{q_{0.75} - q_{0.25}}. \quad (4)$$

208 For  $\alpha = 0.05$ , NP Kurtosis is the difference between the 95th and 5th percentiles, relative  
 209 to the interquartile range. To facilitate interpretation, we scale NP Kurtosis relative to  
 210 its value for Gaussian distribution, and subtract 1. We call this “NP Excess Kurtosis”.  
 211 The value for a Gaussian distribution is zero. A value of 0.2 means that the tails are (on  
 212 average) 20% heavier than those of a Gaussian with the same interquartile range, and  
 213 a value of -0.2 means that the tails are (on average) 20% lighter than a Gaussian with  
 214 the same interquartile range. We calculate NP Kurtosis using  $\alpha = 0.05$  unless otherwise  
 215 stated, to focus on the tail edges, but again this is somewhat arbitrary.

216 Figure 2C,D illustrate how, applied to exactly the same simulated samples, the non-  
 217 parametric measures of skewness and kurtosis produce a smaller fraction of highly in-  
 218 accurate estimates caused by a few extreme values in the sample. But also note that, in  
 219 contrast to the moment-based measures, numerically small values of the NP measures  
 220 (e.g., 0.1 or 0.2) should not be disregarded, because they are both scaled so that a value  
 221 of 1 indicates extremely large departures from a Gaussian distribution.

222 Quantile-based estimation of skewness and kurtosis carries the added value that  
 223 quantile regression methods may be used to derive these properties of size transitions  
 224 as continuous functions of initial size or expected future size. In the examples below,  
 225 we use the **qgam** package to fit smooth additive quantile regression models, which have  
 226 the flexibility to accommodate non-linear size-dependence in skewness and kurtosis.  
 227 One risk of a gam-based approach is that fitted quantiles may be too “wiggly” without  
 228 constraints on their complexity (in the examples below, we specify  $k = 4$  to constrain the

dimension of the basis function). For the gam-averse, other quantile regression models may be equally suitable. For consistency with non-parametric skewness and kurtosis, we similarly use quantile-based measures of mean and variance and quantile regression to visualize these as size-dependent. Specifically, following Wan et al. (2014),

$$\text{NP Mean} = \frac{q_{0.25} + q_{0.5} + q_{0.75}}{3} \quad (5)$$

and

$$\text{NP SD} = \frac{q_{0.75} - q_{0.25}}{1.35}. \quad (6)$$

## 1 Case study: Sea fan corals, *Gorgonia ventalina*

We begin with a simple example where current size is the only predictor of future size. Bruno et al. (2011) developed an IPM to understand the rise and fall of a fungal pathogen *Aspergillus sydowii* in Caribbean sea fan corals *G. ventalina*. The model was based on repeated observations of marked corals in permanent transects at several sites near Akumal, Mexico, recording disease status (infected/uninfected) and the area of uninfected tissue. The epidemic peak had passed and disease incidence was already low, so infected fans were relatively infrequent. We therefore limit the analysis here to uninfected individuals. Bruno et al. (2011) found statistically significant year and site effects, but as those explained a very small fraction of the variation in growth increments, they fitted a single growth model to data pooled across years and sites. We do the same here. The pooled data set consists of 358 observed size transitions. The data exhibited size-dependent variance in growth (change in area,  $cm^2$ ), which Bruno et al. (2011) chose to stabilize by transforming size, using the cube-root of total fan area as the size measure (fig. ??B), and then fitting the standard model with Gaussian growth increments. Here we take a different approach, modeling size-dependent variance explicitly rather than trying to transform it away.

We develop a model using natural log transformation of area. With initial size as the only predictor, a simple way to fit a Gaussian model with nonconstant variance is the `gam` function in `mgcv` library (Wood, 2017) using the `gaulss` family. The mean and standard deviation are both fitted as smoothing spline functions of initial size, and the `predict` function returns the fitted mean and also the inverse of the fitted standard deviations with which we can compute the scaled residuals:

```
# XH is a data frame holding the data
# logarea.t0, .t1 denote initial and final values of log-transformed area
```

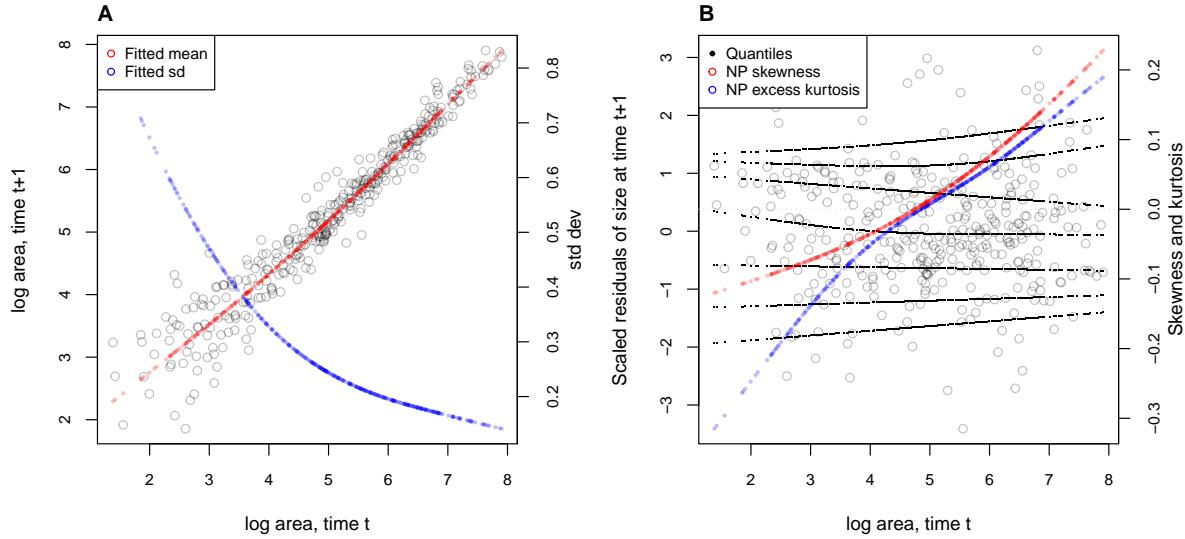


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (blue) derived from them. Black lines in **B** show the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles. Figure made by script AkumalCorals\_qgam.R.

```

261 fitGAU <- gam(list(logarea.t1~ s(logarea.t0), ~ s(logarea.t0)),
262   data=XH, gamma=1.4, family=gaulss())
263 fitted_all = predict(fitGAU,type="response");
264 fitted_sd = 1/fitted_all[,2];
265 scaledResids = residuals(fitGAU,type='response')/fitted_sd;

```

Fig. 3A shows the log-transformed data and Gaussian model. The mean function (solid blue curve) is visually nearly linear, but the fitted nonlinear spline is strongly favored over a linear model for the mean ( $\Delta AIC \approx 9$ ). The spline for standard deviation  $\sigma$  versus initial size shows that smaller individuals exhibit greater variability in future size.

There are no blatant signs of trouble in the pilot Gaussian model, but quantile regressions on the scaled residuals, and the NP Skewness and Kurtosis metrics derived from them (Eq. 3 and 4), suggest deviations from normality (Fig. 3B). Specifically, skewness switches from negative to positive across the size distribution, with smaller corals more likely to shrink than grow and larger corals more likely to grow than shrink. Kurtosis also changes direction over the size distribution, with smaller initial sizes having thinner tails and larger initial sizes having fatter tails than Gaussian. Are these apparent deviations from normality

Consequently, we want to find a better fitting distribution. As noted in the *Introduction* the overall properties of the scaled residuals may be misleading if the growth distribution changes as a function of initial size. Instead, distribution choice should be guided by properties of distributions conditional on initial size. Multiple observations with identical initial sizes will be rare, but we can approximate conditional distributions by grouping size transitions with similar initial sizes. In fig. ??, summary statistics (mean, standard deviation, NP skewness, NP excess kurtosis) are plotted (solid points) for a set of overlapping sliding windows, each containing 10% of the scaled residuals sorted according to initial size. The red dashed lines are what we would see if the pilot model were exactly right: zero mean, skewness and excess kurtosis, and standard deviation equal to one. Black curves are spline regressions through the points using `gam`. Panels A) and B) are a “sanity check” to verify that the pilot model was adequate to capture the mean and variance of growth. Panel C) indicates a shift from negative skew at small sizes to positive skew at larger sizes. Similarly, panel D), reveals that NP kurtosis varies with initial size, negative for small fans and positive at large sizes.

Fitting those properties requires a four-parameter distribution on  $(-\infty, \infty)$  that allows nonzero skew and both positive and negative excess kurtosis. That narrows the options considerably. Of the more than 50 distributions provided in the `gamlss.dist` package, the options are EGB2, JSU, GT, SHASH and four SEP distributions (four different ways of adding skew to the exponential power distribution). We omit SEP2 because it is known to have poorly identified parameters (TJ and AC, 2004), which is problematic for choosing the non-Gaussian model structure in the way we propose.

To choose among the candidate distributions, we sorted the scaled residuals based on initial size into 8 equal-size bins, and used maximum likelihood to fit all candidate distributions to each bin.<sup>3</sup> As all distributions have the same number of parameters, comparison of maximized likelihoods is equivalent to comparing AIC values. With the exception of EGB2 all distributions had very similar maximized likelihoods for all bins; the best overall (summed likelihood) was SEP1, so we proceeded with SEP1. With 5 bins instead of 8 the outcome was the same.

Most four-parameter distributions have the inconvenient property that their parameters are *not* the mean, standard deviation, skew, and kurtosis of the distribution. This happens because (for example) introducing nonzero skew changes the mean and the kurtosis. So to see how distribution parameters vary as a function of initial size, instead

---

<sup>3</sup>Our function `gamlssMaxlik` in online script `fitChosenDists.R` is patterned after the `gamlss` function `gamlssML`, but uses the `maxLik` function from the `maxLik` package (Henningsen and Toomet, 2011), and repeated optimization from different starting values, to increase the reliability of the results.

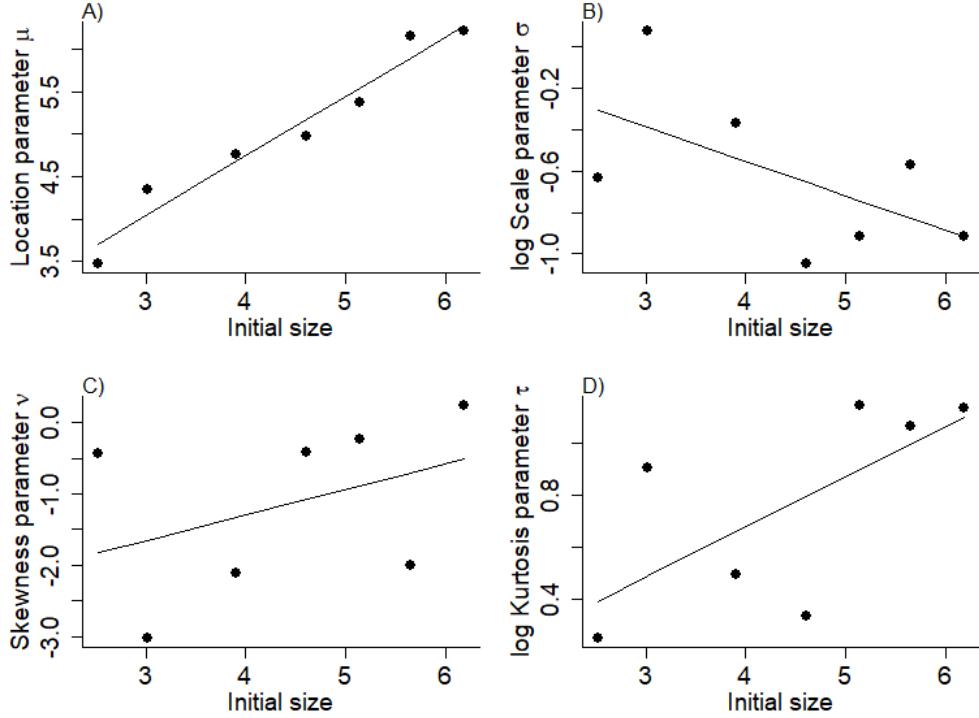


Figure 4: Binned data SEP1 parameter estimates for growth distributions of sea fan corals, *Gorgonia ventalina*. Figure made by scripts `AkumalCorals.R` and `Diagnostics.R`.

311 of inspecting fig. ??, we need to look directly at how parameters vary as a function of  
 312 initial size. As before, we did this by binning the data based on initial sizee, and fitting  
 313 an SEP1 distribution to the subsequent size values in each data subset. Note, in fig. ??  
 314 we used scaled residuals as the response variable, to minimize how variation in mean  
 315 and standard deviation within a bin would affect the skew and kurtosis. Now, because  
 316 all four distribution parameters interact to determine the distribution, we need to fit all  
 317 four simultaneously to the data in each bin.

318 As expected from the pilot Gaussian fit, the location parameter  $\mu$ , the log of the scale  
 319 parameters  $\sigma$ , the skewness parameter  $\nu$  and log of the kurtosis parameter  $\tau$  all appear to  
 320 vary linearly with initial size (log is the default link function for  $\sigma$  and  $\tau$  in SEP1 because  
 321 both must be positive). The corresponding overall growth model is an SEP1 distribution  
 322 in which  $\mu, \log \sigma, \nu$  and  $\log \tau$  are all linear functions of initial size. However, following  
 323 our suggested workflow, we specify  $\mu$  as a quadratic function of initial size because that  
 324 relationship was found to be nonlinear in the pilot Gaussian model. This model is easily  
 325 fitted by maximum likelihood, using the pilot Gaussian fit and the values plotted in fig.  
 326 4 to inform the starting parameter values. The quadratic term for  $\mu$  proved to be highly  
 327 significant ( $p < 0.001$  based on the asymptotic standard error from the inverse Hessian).

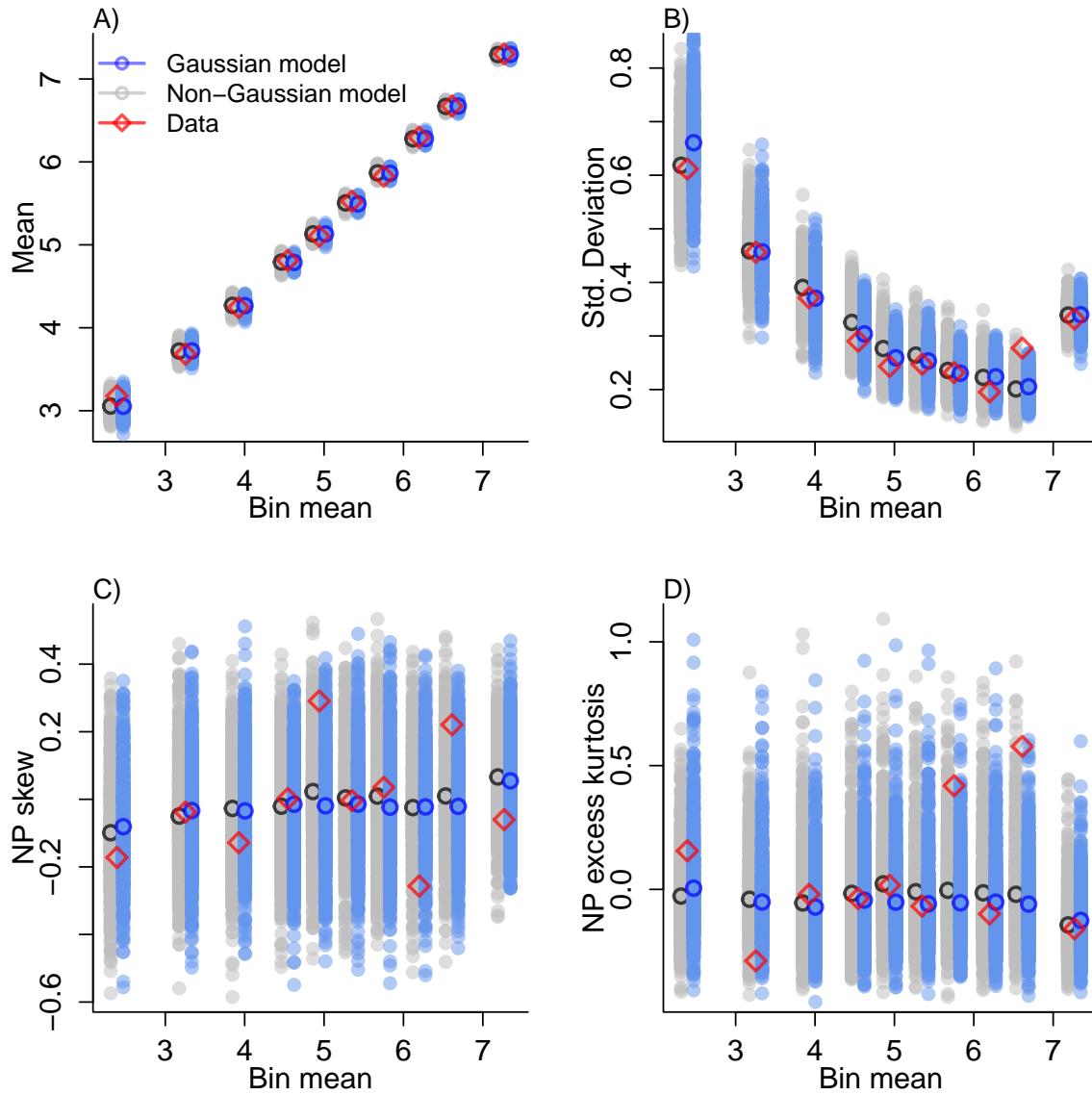


Figure 5: Binned data comparison of moments between 500 simulations of the fitted SEP1 model (grey, black circles), 500 simulations of the fitted Gaussian model (light blue and blue circles), and the actual data (red diamonds) for sea fan corals *G. ventalina*. Grey and light blue circles show the moments for each of 500 replicate simulations, in which the fitted growth models were applied to all initial sizes in the empirical data sets, and black/dark blue circles are the mean across replicates. Red diamonds are values for the data. Bins are defined by initial size. Figure made by scripts Akumal\_Corals.R and Diagnostics.R.

328 The “acid test” of a growth model is that simulations from the fitted model should

329 look like the real data. To implement that criterion, we compared simulated and actual

330 data with respect to moments of the distribution of subsequent size, again using a series

331 of bins defined by initial size to account for how those quantiles vary with initial size. As  
332 a benchmark, we did the same for simulations of the pilot Gaussian model. The results,  
333 in fig. 5, show depressingly small differences between the Gaussian and non-Gaussian  
334 models.

335 What, then, have we gained by fitting a better growth model? Fig. ??A compares  
336 the predicted distributions of subsequent size in the fitted model and Gaussian pilot  
337 models, for the median size of a new recruit (leftmost pair of curves), the median initial  
338 size (central curves), and the 95th percentile of initial size in the data (rightmost curves).  
339 Differences are most pronounced for recruits, which have better odds of growing (lead-  
340 ing to earlier reproduction and higher survival) in the fitted model, small at intermediate  
341 sizes where the skew and excess kurtosis are both small (fig. ??CD), and larger again at  
342 high sizes where the fitted model is leptokurtic. Fig. 6B shows the predicted steady-state  
343 size distributions resulting from a constant unit input of recruits. The main difference  
344 is that the pilot model projects fewer individuals at or near the modal size; overall, the  
345 fitted model predicts 7% more individuals at steady-state. Equivalently, the fitted model  
346 predicts a 7% higher mean lifespan, 19.0 years vs. 17.7 in the pilot Gaussian model.

347 We used `gam` to fit the pilot Gaussian model because that obviated model selection  
348 on the mean and variance functions. However, `gam` should be used with caution. Non-  
349 parametric regression models notoriously “wag their tails” because the ends of the fitted  
350 curve can be pulled close to the outermost data points. This is especially problematic for  
351 growth modeling, because data are typically sparse near the top and bottom of the size  
352 range. To minimize the risk of overfitting we have used `gamma=1.4` to overweight model  
353 degrees of freedom, as suggested by Gu (2013, sec. 3.2). But it is always important to  
354 plot the fitted splines and make sure they do not wag unrealistically. If they do, the  
355 pilot model can be fitted with parametric regression, as we illustrate in our cactus and  
356 creosote bush case studies.

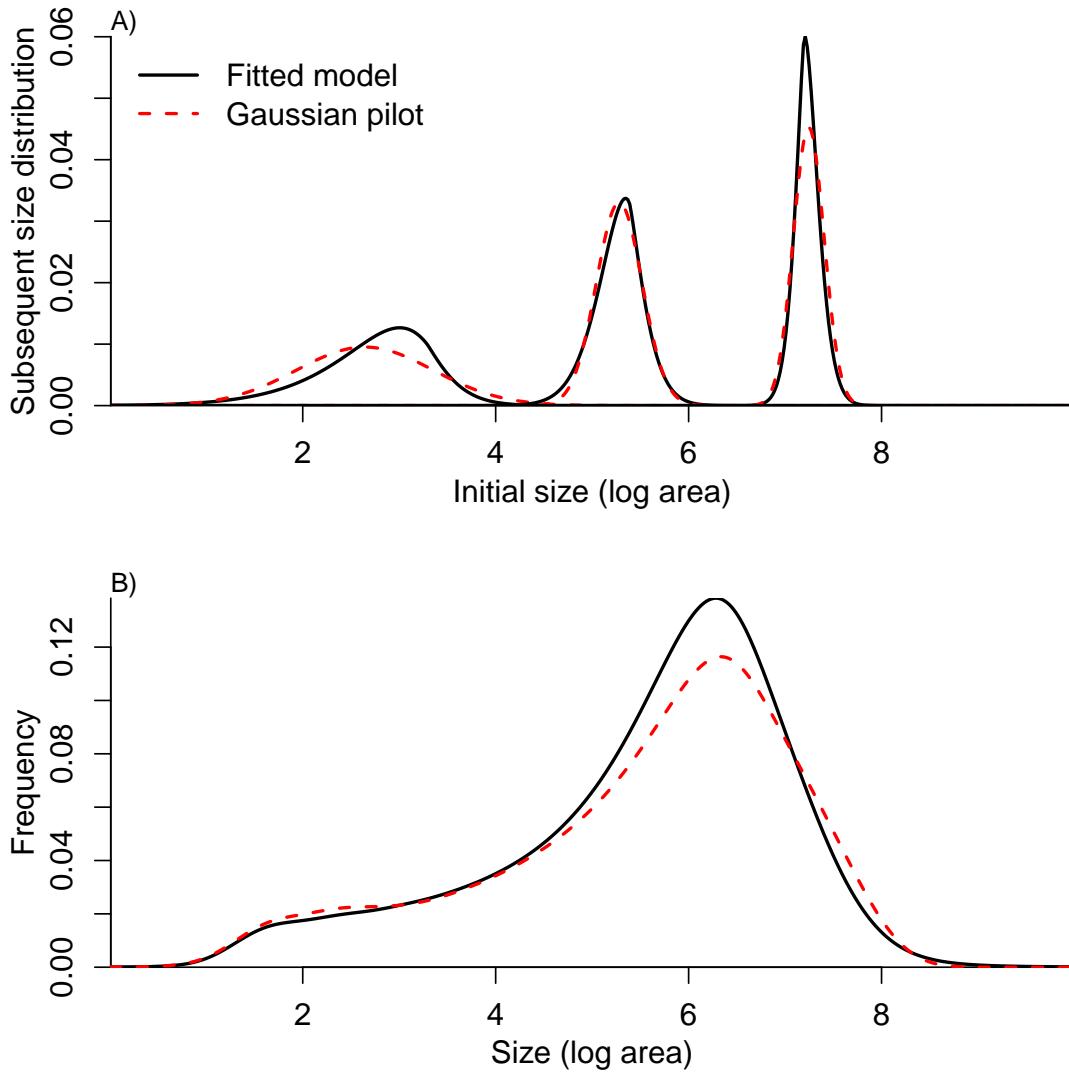


Figure 6: Comparisons between the fitted SEP1 growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year  $t + 1$  for three different values of size in year  $t$ . The leftmost pair of curves are for initial size equal to the median size of a new recruit (data from (Bruno et al., 2011)); central pair is for the median initial size of uninfected individuals; rightmost pair are for the 95th percentile initial size for uninfected individuals. B) Steady-state size distributions resulting from a constant unit input of new recruits. As in Bruno et al. (2011) we assume that larvae are very widely dispersed, so that the number of recruits arriving at any one location is independent of the local population abundance or structure. Size distribution of recruits was described by a kernel density estimate based on the (sadly, only  $n = 9$ ) measured sizes of known new recruits. Figure made by script AkumalCoralsIPMs.R.

## <sup>357</sup> 2 Case study: cactus, *Cylindriopuntia imbricata*

<sup>358</sup> The next case study, focusing on the tree cholla cactus *Cylindriopuntia imbricata* at the  
<sup>359</sup> Sevilleta Long-Term Ecological Research site in central New Mexico, adds an impor-  
<sup>360</sup> tant new feature on top of the simple size-dependent regressions in the previous study:  
<sup>361</sup> random effects associated with temporal (year) and spatial (plot) environmental hetero-  
<sup>362</sup> geneity. This long-term study of cactus demography was initiated in 2004 and different  
<sup>363</sup> subsets of the data have been analyzed in various IPM studies, all using Gaussian growth  
<sup>364</sup> kernels (Compagnoni et al., 2016; Czachura and Miller, 2020; Elderd and Miller, 2016;  
<sup>365</sup> Miller et al., 2009; Ohm and Miller, 2014). In fact, (Elderd and Miller, 2016) presented a  
<sup>366</sup> Gaussian growth model fit to the cactus data as an example of a well fit growth function,  
<sup>367</sup> based on a marginal distribution of residuals that appeared approximately Gaussian  
<sup>368</sup> and posterior predictive checks (PPCs) of a Bayesian model that suggested consistency  
<sup>369</sup> between the real data and data simulated from the fitted model (Fig. 4 in (Elderd and  
<sup>370</sup> Miller, 2016)). While PPCs and the associated “Bayesian P-value” are popular diagnostic  
<sup>371</sup> tools, they are often considered to be too conservative (Conn et al., 2018; Zhang, 2014),  
<sup>372</sup> failing to reject marginally bad models even though they are very effective in rejecting  
<sup>373</sup> models that are terrible. The choice of discrepancy function (the statistic used to de-  
<sup>374</sup> scribe real and simulated data) can also be limiting: in our previous work, we used a  
<sup>375</sup> discrepancy function focused on variance (the sum of the squared residuals), so we had  
<sup>376</sup> a built-in blind-spot for mismatches in higher moments. In the clarity of hindsight, the  
<sup>377</sup> PPC gave a false sense of security; the Gaussian was a poor choice all along.

<sup>378</sup> The data for this new analysis include 5203 size transition observations from 924  
<sup>379</sup> unique individuals spanning nine transition years (2009–2017) and eight 30m-by-30m  
<sup>380</sup> plots (we excluded earlier years of data corresponding to a different cohort of plants  
<sup>381</sup> from a different set of locations). Following previous studies, we quantified size as the  
<sup>382</sup> natural logarithm of plant volume ( $cm^3$ ), derived from height and width measurements.

### <sup>383</sup> 2.1 A pilot Gaussian model

<sup>384</sup> We begin the growth modeling workflow with a simple mixed model for size in year  $t + 1$   
<sup>385</sup> as a function of size in year  $t$ , with random intercepts for year and plot and assuming  
<sup>386</sup> normally distributed residuals. We considered four candidate models, with and without  
<sup>387</sup> a quadratic term in the predictor for the mean and with residual variance as a function  
<sup>388</sup> of either initial size or the model’s fitted value (which includes initial size plus the  
<sup>389</sup> random effects). We allowed for non-constant variance by iterively re-fitting the models

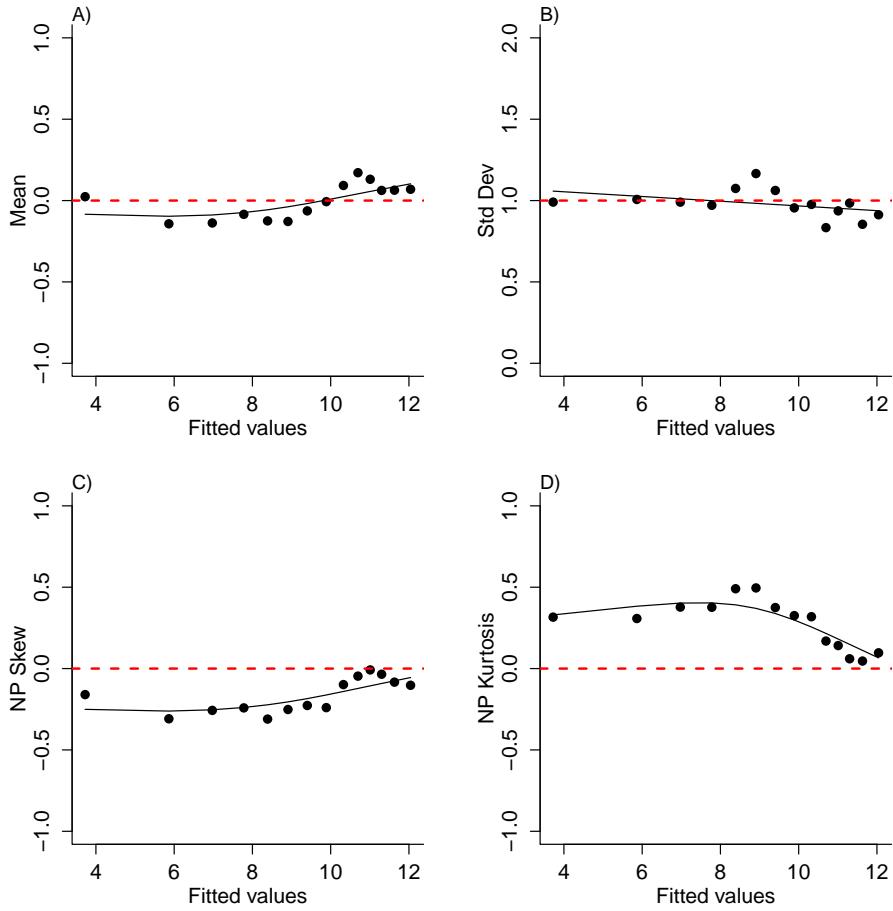


Figure 7: Rolling-moment analysis of standardized residual from a Gaussian fit to the cactus growth data. Red dashed lines indicate the Gaussian expectation. These data show skewness and kurtosis that deviate from Gaussian.

390 with weights based on a second-order polynomial that related the *SD* of the residuals  
 391 to initial size or fitted values. We iterated until there was effectively no change in the  
 392 model weights, indicating convergence on maximum likelihood parameter estimates for  
 393 the coefficients of the mean and *SD* linear predictors.

394 The best-fit model from this pilot analysis included the quadratic term for the mean  
 395 and residual variance as a function of initial size. However, while the standardized  
 396 residuals from this model are approximatley mean zero and unit variance (as they should  
 397 be), there are clear deviations from normality in the higher moments: negative skew  
 398 (decreases in size were more common than increases, visible in the raw data [Fig. ??])  
 399 and positive excess kurtosis, both greater at smaller sizes and minimal at larger sizes  
 400 (Fig. 7).

## 401 2.2 An improved growth model

402 To better capture size transitions, we need a distribution with negative skew and positive  
403 excess kurtosis, but both of which may be negligible at some sizes. Appropriate candi-  
404 dates include Johnson's  $S_U$  distribution, which is limited to positive excess kurtosis, and  
405 the sinh-arcsinh (SHASH) and skewed power exponential distributions, which can vary  
406 from leptokurtic to platykurtic (Fig. 1). As above, we divided the standardized residuals  
407 into discrete bins of initial size and fit competing distributions to each data subset using  
408 our function `gamlssMaxlik()`. We found that Johnson's  $S_U$  distribution (JSU in **gamlss**)  
409 was favored over most of the size distribution, particularly the larger sizes. However, at  
410 smaller sizes, the SHASH distribution was favored. They are both four-parameter dis-  
411 tributions but the SHASH is more flexible than the JSU, capable of capturing a greater  
412 range of possible kurtosis for a given amount of skew<sup>4</sup>. We ultimately settled on the  
413 SHASH distribution, but only after first trying the JSU and proceeding through the  
414 workflow (Fig. 1). We were unsatisfied with the correspondence between real and simu-  
415 lated data in the final step, so we followed the re-iteration loop back to "choose a better  
416 distribution", this time the SHASH. To keep this section concise, we do not fully nar-  
417 rate that re-iteration loop, but we expect it will be a common feature of many growth  
418 analyses.

419 To guide the final fit, we visualized parameter estimates of the SHASH distribution  
420 across bins of initial size, which suggested second-order polynomials for the parameters  
421 that control variance, skew, and kurtosis (Fig. 8). The final likelihood model was thus:

```
422 LogLik=function(pars,response,U){  
423   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];  
424   mu = U%*%pars1;  
425   val = dJSU(x = response,  
426               mu=mu,  
427               sigma = exp(pars2[1] + pars2[2]*x + pars2[3]*x^2),  
428               nu = pars2[4] + pars2[5]*x + pars2[3]*x^2,  
429               tau = exp(pars2[6] + pars2[7]*x + pars2[8]*x^2), log=T)  
430   return(val)  
431 }
```

---

<sup>4</sup>This comes from Steve's *NPSkewKurtosisRanges.pdf*

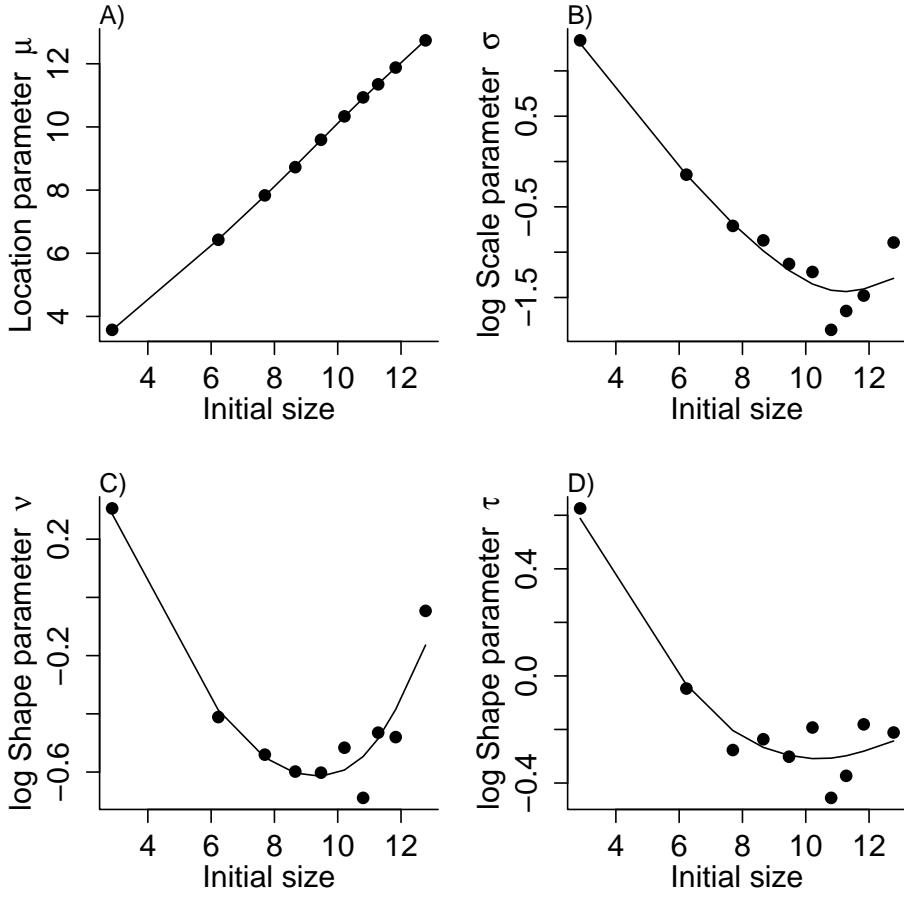


Figure 8: Parameter estimates of the SHASH distributions fit to cactus growth data in discrete bins of initial size.

432 Here, response is future size and  $U$  is the model matrix for the location parameter of  
 433 the SHASH, derived from the mean of the best-fit Gaussian model, but now converting  
 434 random effects of years and plots to fixed effects:

435 `U=model.matrix(~ 0 + year_t + plot + log(vol_t)+ I(log(vol_t)^2), data=CYIM)`

436 We can use the ‘shrinkage’ methods described in Appendix S.1 to estimate the vari-  
 437 ance terms associated with year and plot effects, even though they are fit here as fixed  
 438 effects. There is a tedious but not insurmountable complication that can arise in models  
 439 with multiple random effects, as in this one. By supressing the intercept in the model  
 440 matrix (the zero after the tilde), we get parameter estimates for each year rather than  
 441 offsets relative to a baseline year. However, because there is the additional effect of plot,  
 442 all of the year effects are conditioned on the first level of the plot factor variable (plot 1)  
 443 and the remaining plot coefficients are offsets of plot 1. The parameter estimate for plot  
 444 1 is therefore the mean of the year estimates, but there is no way to calculate its standard

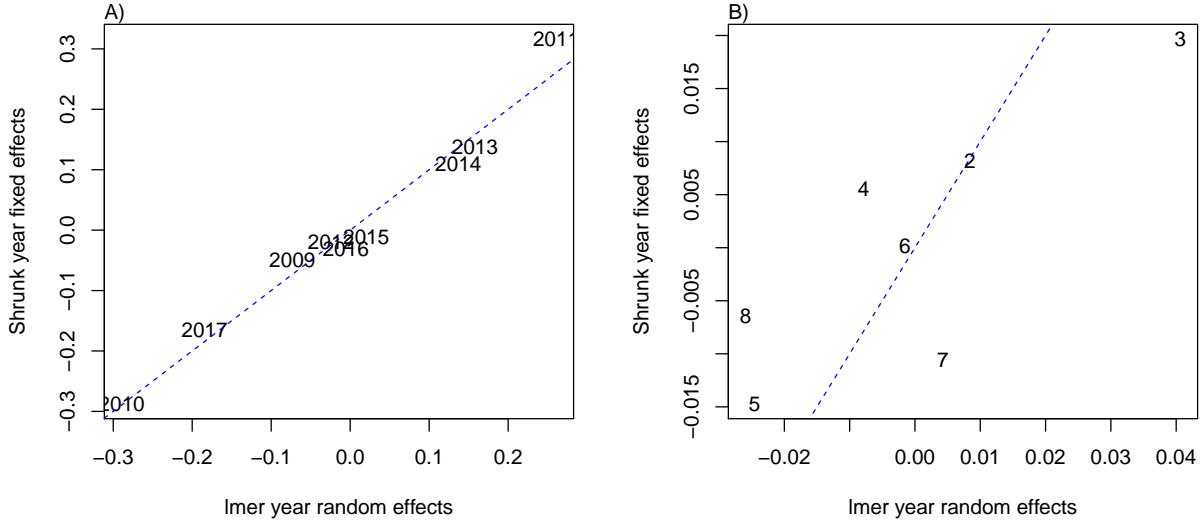


Figure 9: Comparison of year (A) and plot (B) random effects from ‘shrinking’ fixed-effect estimates (y-axes) vs. `lme4` mixed models (x-axes) for cactus data set.

445 error. We estimate the plot variance **excluding plot 1**<sup>5</sup>, a possible source of bias the vari-  
 446 ance estimate. The random effects estimated via the shrinkage method correspond well  
 447 with those from the original Gaussian mixed model, particularly for year effects (Fig.  
 448 9A). Plot effects were less tightly correlated but there was also less variance across plots  
 449 than there was across years (Fig. 9B).

450 Simulations of the final SHASH model – the final step of the workflow – indicate  
 451 that it describes the cactus growth data well (Fig. 10). The SHASH model did not  
 452 provide a noticeable improvement over the Gaussian model in terms of mean and vari-  
 453 ance of size transitions (Fig. 10A,B). This is expected, since an appropriately specified  
 454 Gaussian model with non-constant variance should be able to accommodate this type of  
 455 complexity. It is in skewness and kurtosis that the SHASH really shines (Fig. 10C,D),  
 456 effectively capturing observed trends in these features of size transitions.

---

<sup>5</sup>A very unsatisfying feature of this analysis. I’ve tried flipping the order so that plots are conditioned on year 1, but when I do it this way `sigma2.hat` for plot is negative. Not totally unreasonable because the plot effects are small relative to year effects, but still annoying. Steve has mentioned that we might get around this by abandoning the `model.matrix` function but I don’t see how. We cannot get unconditional estimates of plot and year effects without adding a parameter to the model, and that parameter is unidentifiable.

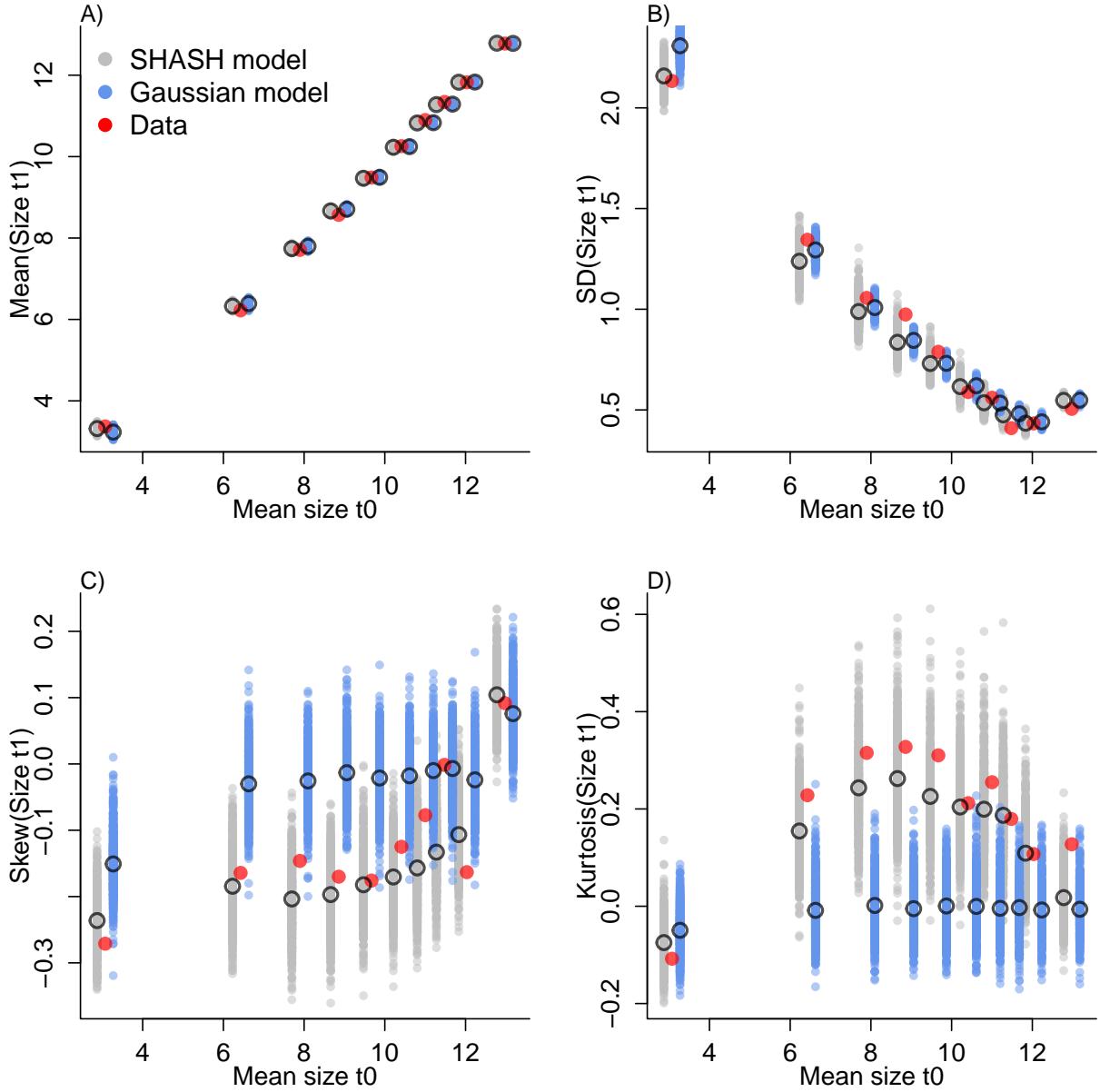


Figure 10: Simulated data from SHASH and Gaussian growth kernels compared to cactus growth data to which the models were fit.

### 457 2.3 Consequences of improved growth modeling for IPM predictions

458 Finally, we can ask how the improvements we have made to modeling cactus growth af-  
 459 fect inferences from the full IPM. Details of IPM construction and analysis are provided  
 460 in Appendix #<sup>6</sup>. Our previous IPM analyses of this system have consistently shown  
 461 asymptotic population growth rates below replacement-level ( $\lambda < 1$ ). Our new results

---

<sup>6</sup>Need to do. Maybe we can have one appendix for all the IPMs.

<sup>462</sup> with the improved SHASH growth kernel are qualitatively consistent with these results.  
<sup>463</sup> However the SHASH growth kernel predicts a much faster rate of decline – a 9% differ-  
<sup>464</sup> ence in annual change – than the Gaussian, holding all else equal ( $\lambda_{SHASH} = 0.901$  vs.  
<sup>465</sup>  $\lambda_{Gaussian} = 0.995$ ). The SHASH also predicts a very different and much smaller stable  
<sup>466</sup> size distribution (SSD) than does the Gaussian (Fig. 11). The observed size distribu-  
<sup>467</sup> tion falls in between the two predictions, suggesting that, either way, this population  
<sup>468</sup> is not at its stable distribution. The difference in predictions of the two IPMs is driven  
<sup>469</sup> by contrasting size transitions of large plants (Fig. 12). This contrast is driven by skew  
<sup>470</sup> and kurtosis, since the mean and variance of the SHASH and Gaussian models were  
<sup>471</sup> nearly identical for large sizes (Fig. 10). The Gaussian growth kernel over-predicts size  
<sup>472</sup> transitions at large sizes, and this leads to the larger median size at SSD and also the  
<sup>473</sup> smaller peak corresponding to new recruits, since reproductive output is strongly size-  
<sup>474</sup> dependent. This peak disappears from the SHASH SSD because reproductive output is  
<sup>475</sup> strongly size dependent, and the SHASH growth kernel predicts that transitions to the  
<sup>476</sup> largest sizes are less likely.

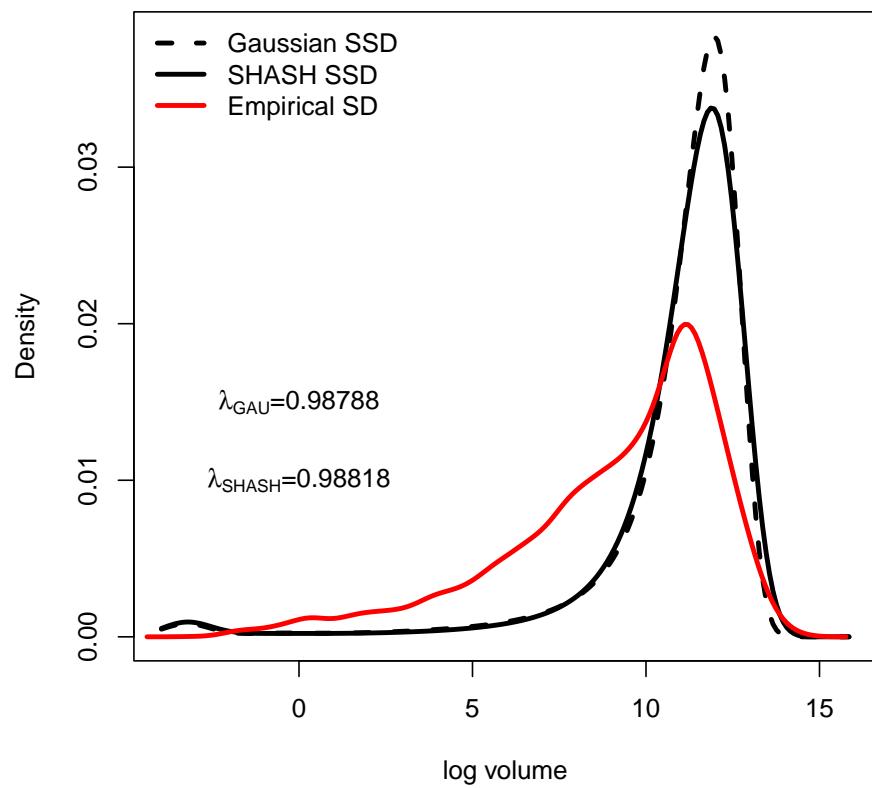


Figure 11: Predicted and observed size distributions of tree cholla cactus.

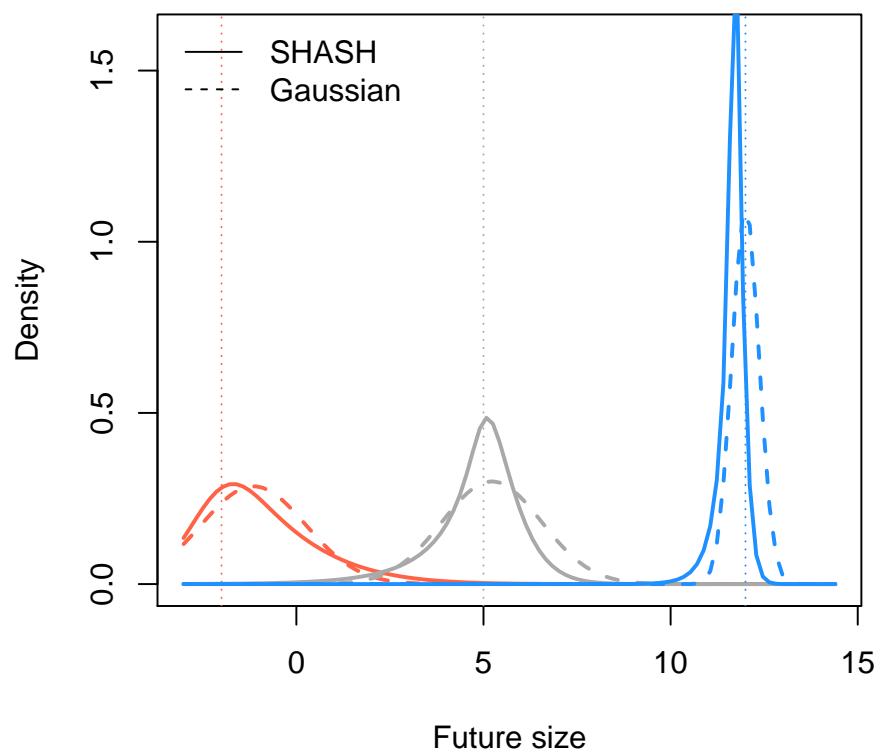


Figure 12: Predicted probability of future size given three initial sizes indicated by vertical lines.

477 **3 Case study: bunchgrass, *Pseudoroegneria spicata***

478 We again consider a species where one of us is the offender – initially using the default  
479 model because it was hard to do better at the time (Adler et al., 2010), but sticking with  
480 it (e.g., Adler et al., 2018; Tredennick et al., 2018) when it would have been easy to do  
481 better.

482 We used the most recently curated version of the data (Adler et al., 2018, at  
483 doi.org/10.5061/dryad.96dn293), both legacy (22 annual transitions between 1926 and  
484 1957) and modern (8 annual transitions from 2008 to 2016, excluding moisture manip-  
485 ulation treatments). We excluded seedlings, which require separate models (Chu and  
486 Adler, 2014, 2015; Snyder and Ellner, 2018), and individuals mapped as “too small to  
487 measure” that should be modeled separately as a discrete size category (though in the  
488 past we have not done that). The measure of plant size was log basal cover.

489 Based on past analyses, (1) we did not distinguish between historical and modern  
490 Control treatments (Adler et al., 2018) (2) we included size by year interactions with  
491 year-specific slope and intercept for a linear relationship between current and future  
492 size (log basal cover); (3) Quadrat group (labeling sets of spatially nearby quadrats),  
493 Treatment (Control or Shrub removal) and competition with other species were included  
494 as covariates. As in past models, competition was measured by distance-weighted cover  
495 of competing species, using nonparametric competition kernels estimated from the data  
496 (Teller et al., 2016).

497 Results:

- 498 1. The Gaussian pilot model was fitted using `gam`, with spline terms for the effect of  
499 initial size on the mean subsequent size, and for the standard deviation. Based on  
500 past analyses, (1) the measure of size was log basal cover; (2) We did not distinguish  
501 between historical and contemporary Control treatments (Adler et al., 2018); (3) the  
502 model included size by year interactions, with year-specific slope and intercept for a  
503 linear effect of current size on future size; (3) quadrat group (labeling sets of spatially  
504 nearby quadrats), Treatment (Control or Shrub Removal) and competition with other  
505 species were included as covariates. As in past models, competition was measured  
506 by distance-weighted cover of competing species, using nonparametric competition  
507 kernels estimated from the data (Teller et al., 2016).
- 508 2. Model selection for the pilot Gaussian model was thus limited to (1) reducing the  
509 number of competition terms by combining or dropping species, and (2) choosing  
510 between two options for the variance: dependence on initial size, or on the predicted

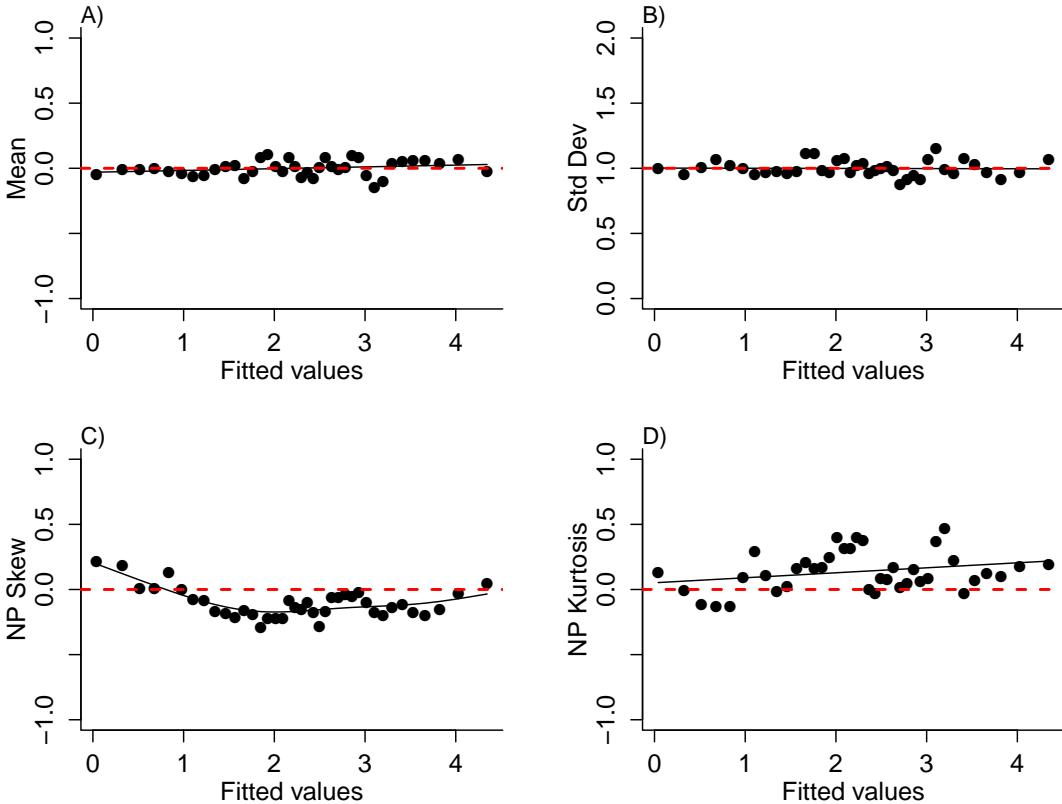


Figure 13: Rolling moments of scaled residuals from the pilot model, as a function of fitted values, for *P. spicata*. The dashed red lines are the value expected if the pilot model fits the data.

mean subsequent size. The latter option is not directly available in `gam`, but was done through iterative re-fitting. Model selection was based on AIC values reported by `gam`). Growth variance depending on the linear predictor was strongly favored ( $\Delta \text{AIC} \approx 50$ ). The selected competition model had three competition covariates: (1) cover of the shrub *Artemisia tripartita*, (2) cover of the other two dominant bunchgrasses combined, and (3) cover of all other species combined.

3. The set of all scaled residuals was non-Gaussian, based on quantile-quantile plot, mostly in the lower tail. Statistical tests confirmed that the standardized residuals are non-Gaussian (all  $P < 0.001$ ).
4. Rolling moments diagnostics on the scaled residuals (Fig. 13) confirm that the mean and SD are nearly constant as a function of the fitted value, as they should be. The small trend in the mean shows that the regression coefficients are slightly biased, which is not unexpected because the Gaussian assumption is violated. The trends in skew and kurtosis are too big to ignore.

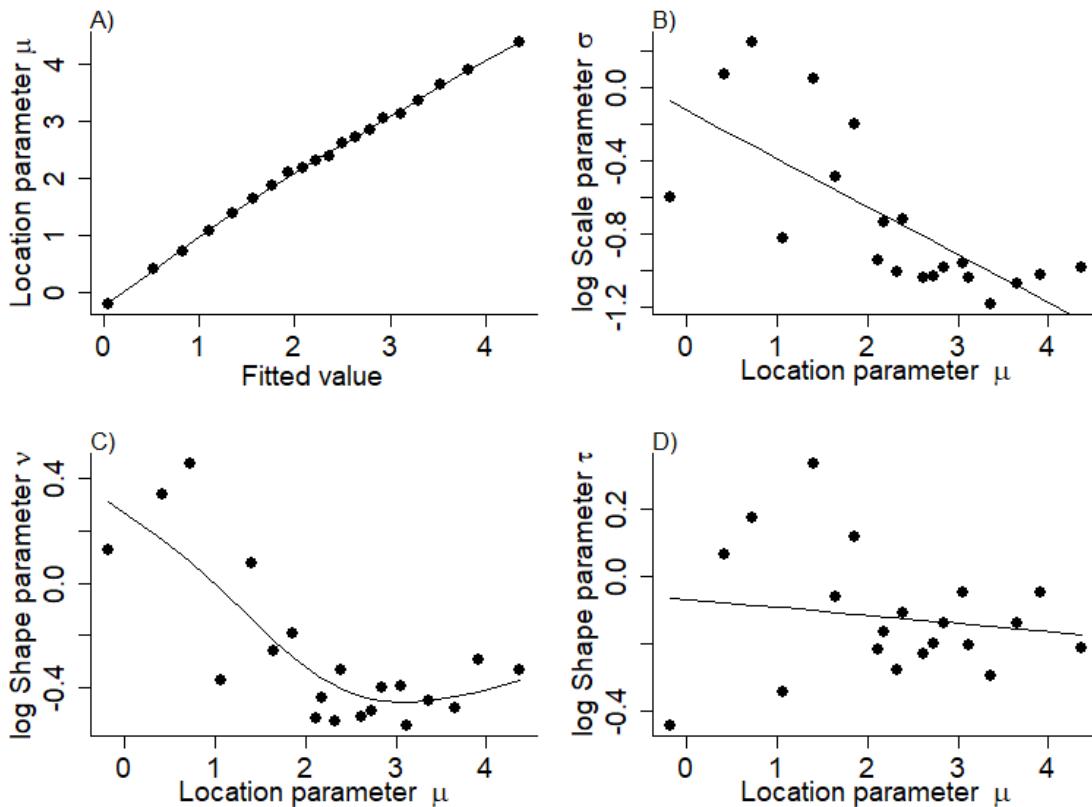


Figure 14: Binned data SHASH parameters plot for *P. spicata*.

- 525 5. What distribution families can accomodate the features in Fig. 13? We need at least  
 526 4 parameters (to allow skew and excess kurtosis), and both postive and zero excess  
 527 kurtosis must be possible. In JSU and skew  $t$  and that makes fitting problematic,  
 528 because zero excess kurtosis only occurs as a limit, not at any actually possible  
 529 parameter values ( $df \rightarrow \infty$  for  $t$ , and  $\tau \rightarrow 0$  for JSU). Comparison was therefore  
 530 limited to SHASH, SHASHo, SEP1, SEP3, SEP4 – not SEP2 for the reasons noted above.  
 531 The winner was SHASH.
- 532 6. Now we need to model the size-dependence of the SHASH distribution parameters.  
 533 In the pilot model,  $\sigma$  as a function of fitted value was strongly favored or  $\sigma$  as a  
 534 function of initial size. We retain that structure in the SHASH model. The data were  
 535 therefore binned based on fitted values of the pilot model, and a SHASH distribution  
 536 was fitted to each bin. fig. 14. Panel A),  $\mu$  as a function of the fitted values, doesn't  
 537 really inform the modeling, but it shows that it is safe to regard  $\mu$  as equivalent to the  
 538 "fitted value" in the pilot model. The other plots tell us how the other distribution

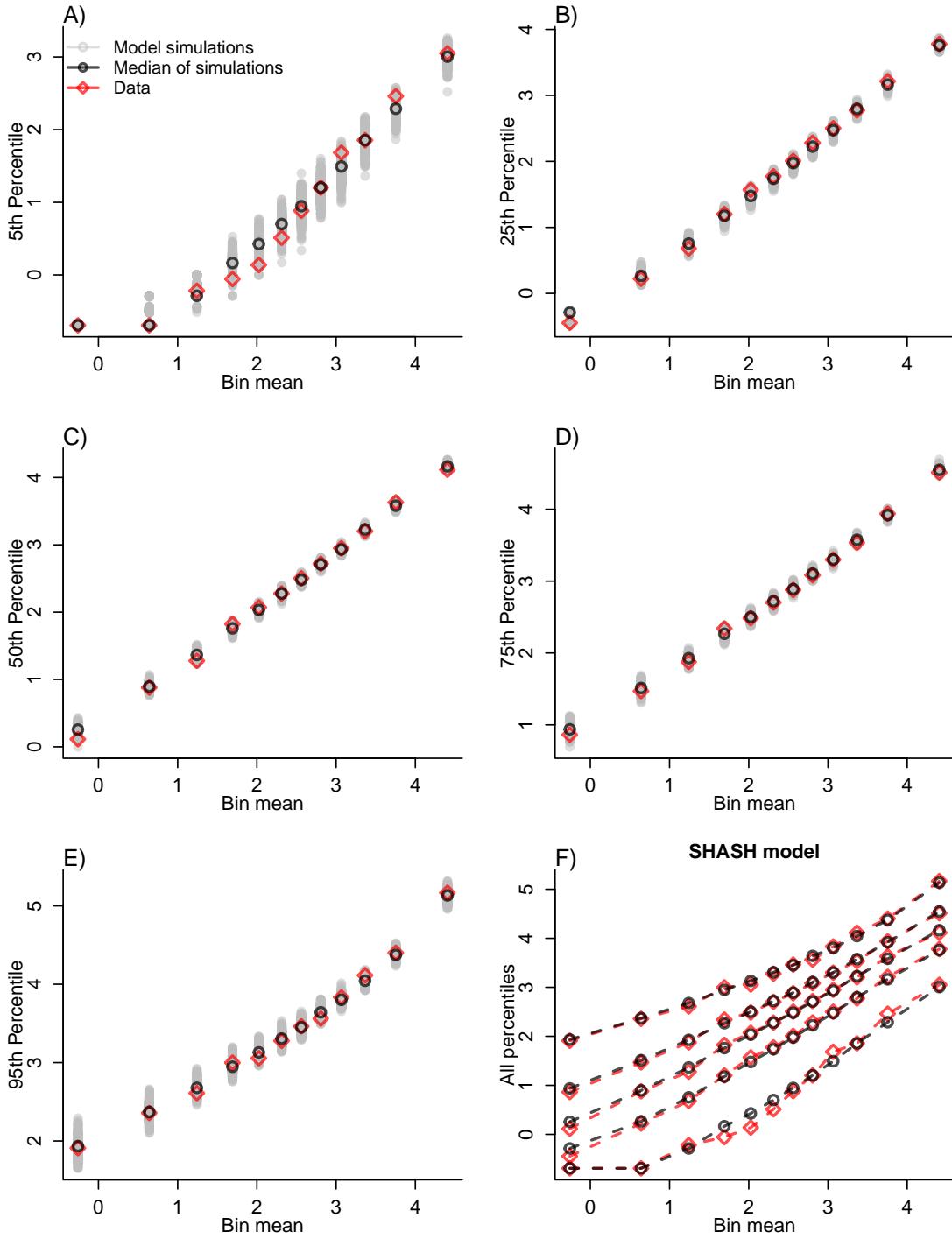


Figure 15: Binned data comparison of distribution quantiles between simulations of the fitted SHASH model (grey, black) and the actual data (red) for *P. spicata*. Individuals were binned based on their initial size.

parameters should be modeled as a function of  $\mu$ .  $\log \tau$  is quadratic, and  $\log v$  is linear.

539

540

Figure 16: Binned data comparison of moments between simulations of the fitted SHASH model (grey, black) and the actual data (red) for *P. spicata*. Individuals were binned based on their initial size.

- 541 7. The model is easy to fit by ML. Does it do a good job of fitting the data? The  
542 comparison based on binned quantiles (fig. 15 looks pretty good except that the  
543 variation in the 5th percentile of the data is more wiggly than the 5th percentile  
544 of the model. To eliminate this imperfection, we could perhaps try higher-order  
545 polynomials for  $\nu$  or  $\tau$ .
- 546 8. More importantly, the SHASH model is a substantial improvement over the pilot  
547 Gaussian model, as seen in the binned moments diagnostic plot, fig. 16. The fits  
548 to mean and standard deviation are about as good, but only the SHASH model  
549 captures the skew and kurtosis and how they vary with size.
- 550 9. Finally, a simulation study was done to see how well the shrinkage approach re-  
551 covers known year effects. Fig. 17 illustrates the results for the intercept coefficient.  
552 Panels A,B,C,D are the fixed-effects estimates, recommended shrunk estimates, more  
553 strongly shrunk estimates, and estimates from lmer using the fitted standard devia-  
554 tion from the pilot Gaussian model. Except for lmer they all look very similar, and  
555 they are. The recommended shrunk estimates had the lowest mean-square error for  
556 both slope and intercept (averaged across all years), and their sample variance was  
557 closest to the across-year variance in the data-generating model, but the improve-  
558 ment over the fixed-effects estimates was small:

559 Intercept: true SD = 0.497, fixed SD = 0.516, shrunk SD = 0.468, lmer SD = 0.418

560 Slope: True SD = 0.129, fixed SD = 0.137, shrunk SD = 0.126, lmer SD = 0.106

561 Note that for both slope and intercept, the lmer estimates were over-smoothed (they  
562 under-estimated the actual between-year variance).

563 This study aligns with our earlier finding for IPMs with temporal random effects  
564 (Metcalf et al., 2015): the simplest method (fitting as fixed effects) works about as  
565 well as anything else. In that study shrinkage was not helpful at all, except that the  
566 fixed-effects and shrunk estimates were biased in opposite directions and therefore  
567 were useful as bounds on the truth. Here, shrinkage makes a small improvement.  
568 However, the biggest improvement in random effects estimation comes from using  
569 a better, non-Gaussian model for the growth distribution.

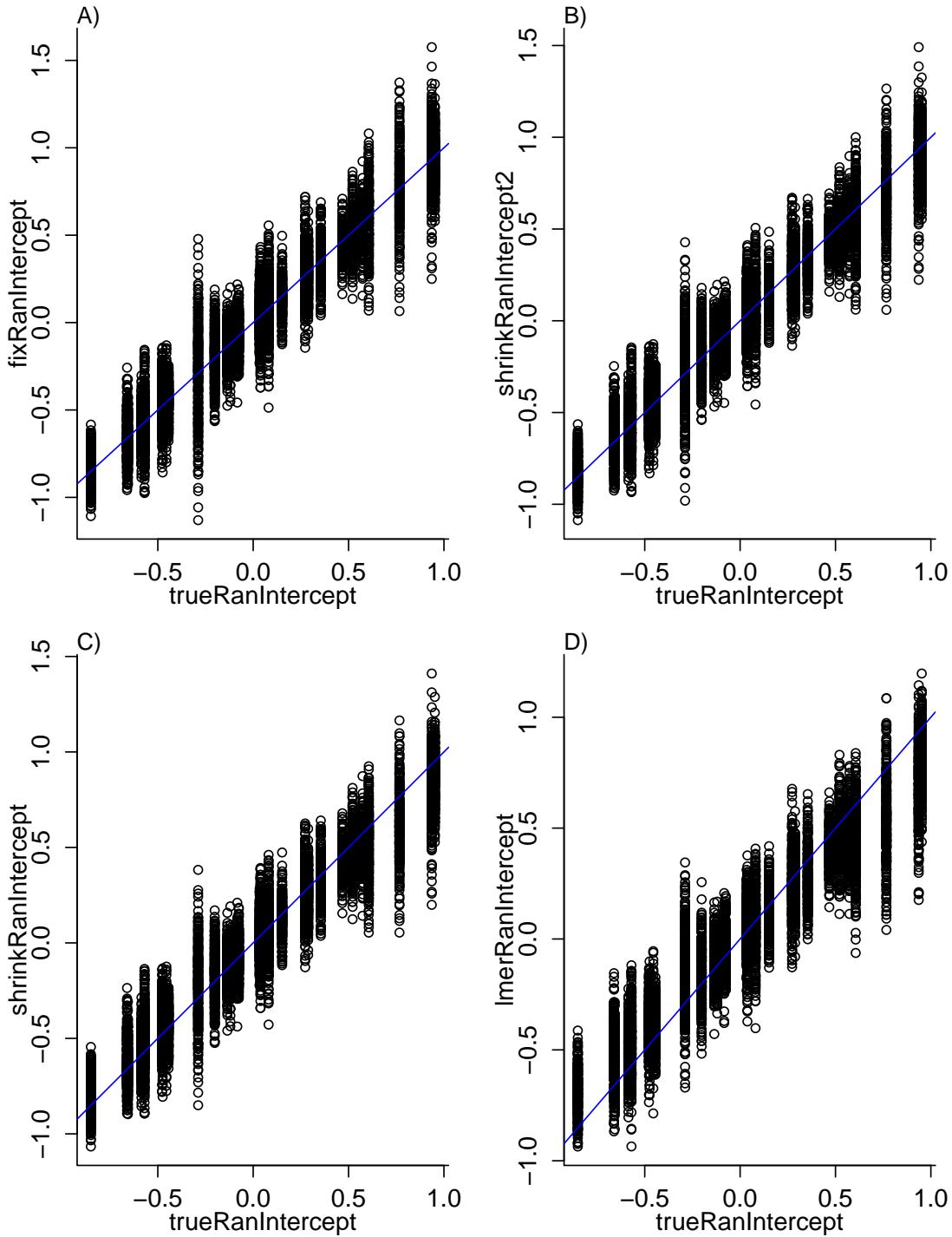


Figure 17: Comparison of “true” and estimated intercept year effects for *P. spicata*.

570 **4 Case study: creosote bush, *Larrea tridentata***

571 Our final case study comes from another perennial plant, the woody shrub *Larrea triden-*  
572 *tata*, whose dynamics we are studying at the Sevilleta LTER. At this site as elsewhere in  
573 the Southwest US, creosote bush is invading and displacing desert grassland habitats ().  
574 The data described here were collected to understand patterns of density dependence in  
575 creosote bush demography, asking whether vital rates are maximized approaching zero  
576 density, at the leading edge of the expansion front (consistent with a ‘pulled’ invasion),  
577 or whether there is a demographic advantage for shrubs at higher densities due to pos-  
578 itive feedbacks expected for some ecosystem engineers (leading to a ‘pushed’ invasion).  
579 We step through the growth analysis following our suggested workflow and then con-  
580 nect the growth model to a spatial integral projection (SIPM) model that allows us to  
581 ask how ‘improved’ growth modeling changes predictions for the speed of creosotebush  
582 encroachment (T. Drees, B. Ochocki, S. Collins, T. Miller, *ms in prep*). The key features of  
583 this growth analysis are: initial size-dependence

584 **4.1 A pilot Gaussian model**

585 **5 Discussion**

586 Here are some of the issues to be discussed.

- 587 • Modeling the mean with gam vs glm
- 588 • Modeling variance and higher moments as functions of covariates vs fitted values
- 589 • Choosing a better distribution – how to make the choice
- 590 • Comparison of our method with beta regression
- 591 • We have emphasize growth but same principles apply to other continuous state  
592 transitions, eg disease IPMs.

593 **Acknowledgements**

594 This research was supported by US NSF grants DEB-1933497 (SPE) and ....

595 **6 Authorship statement**

596 All authors discussed all aspects of the research and contributed to developing methods,  
597 analyzing data, and writing and revising the paper.

598 **7 Data accessibility statement**

599 No original data appear in this paper. Should the paper be accepted, all computer scripts  
600 supporting the results will be archived in a Zenodo package, with the DOI included at  
601 the end of the article. During peer review, our data and code are available at [https://github.com/texmiller/IPM\\_size\\_transitions](https://github.com/texmiller/IPM_size_transitions).  
602

603 **Literature Cited**

- 604 Adler, P. B., Ellner, S. P., and Levine, J. M. (2010). Coexistence of perennial plants: an  
605 embarrassment of niches. *Ecology Letters*, 13(8):1019–1029.
- 606 Adler, P. B., Kleinhesselink, A., Hooker, G., Taylor, J. B., Teller, B., and Ellner, S. P.  
607 (2018). Weak interspecific interactions in a sagebrush steppe? Conflicting evidence  
608 from observations and experiments. *Ecology*, 99(7):1621–1632.
- 609 Anscombe, F. J. and Glynn, W. J. (1983). Distribution of the kurtosis statistic  $b_2$  for  
610 normal samples. *Biometrika*, 70(1):227–234.
- 611 Bates, D., Sarkar, D., Bates, M. D., and Matrix, L. (2007). The lme4 package. *R package  
612 version*, 2(1):74.
- 613 Bruno, J. F., Ellner, S. P., Vu, I., Kim, K., and Harvell, C. D. (2011). Impacts of aspergillosis  
614 on sea fan coral demography: modeling a moving target. *Ecological Monographs*,  
615 81(1):123–139.
- 616 Chu, C. and Adler, P. B. (2014). When should plant population models include age  
617 structure? *Journal of Ecology*, 102(2):531–543.
- 618 Chu, C. and Adler, P. B. (2015). Large niche differences emerge at the recruitment stage  
619 to stabilize grassland coexistence. *Ecological Monographs*, 85(3):373–392.
- 620 Compagnoni, A., Bibian, A. J., Ochocki, B. M., Rogers, H. S., Schultz, E. L., Sneck, M. E.,  
621 Elderd, B. D., Iler, A. M., Inouye, D. W., Jacquemyn, H., et al. (2016). The effect of  
622 demographic correlations on the stochastic population dynamics of perennial plants.  
623 *Ecological Monographs*, 86(4):480–494.
- 624 Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide  
625 to bayesian model checking for ecologists. *Ecological Monographs*, 88(4):526–542.
- 626 Cooch, E. G. and White, G. C. (2020, accessed 5/17/2020). *Program MARK - a 'gentle  
627 introduction'*. Available at phidot.org.
- 628 Coulson, T. (2012). Integral projections models, their construction and use in posing  
629 hypotheses in ecology. *Oikos*, 121(9):1337–1350.
- 630 Crone, E. E. (2016). Contrasting effects of spatial heterogeneity and environmental  
631 stochasticity on population dynamics of a perennial wildflower. *Journal of Ecology*,  
632 104(2):281–291.

- 633 Czachura, K. and Miller, T. E. (2020). Demographic back-casting reveals that subtle  
 634 dimensions of climate change have strong effects on population viability. *Journal of*  
 635 *Ecology*.
- 636 D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1.  
 637 *Biometrika*, pages 679–681.
- 638 Davis, C. (2015). *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.
- 639 Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying  
 640 a new structured population model. *Ecology*, 81(3):694–708.
- 641 Elderd, B. D. and Miller, T. E. (2016). Quantifying demographic uncertainty: Bayesian  
 642 methods for integral projection models. *Ecological Monographs*, 86(1):125–144.
- 643 Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E., and Rees, M. (2022).  
 644 A critical comparison of integral projection and matrix projection models for demo-  
 645 graphic analysis: Comment. *Ecology*, 103(10):e3605.
- 646 Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modeling of Structured Popula-*  
 647 *tions: A Practical Guide to the Integral Projection Model*. Springer, New York.
- 648 Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in  
 649 animal populations. *Ecology*, 79:2531 – 2538.
- 650 Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New  
 651 York, 2 edition.
- 652 Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed  
 653 models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.
- 654 Henningsen, A. and Toomet, O. (2011). maxlik: A package for maximum likelihood  
 655 estimation in r. *Computational Statistics*.
- 656 Hérault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C. T., Wagner,  
 657 F., and Baraloto, C. (2011). Functional traits shape ontogenetic growth trajectories of  
 658 rain forest tree species. *Journal of ecology*, 99(6):1431–1440.
- 659 Jones, M. and Pewsey, A. (2009). *Biometrika*, 96:761 – 780.
- 660 Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis.  
 661 *The American Statistician*, 65(2):89 – 95.

- 662 Komsta, L. and Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and  
663 related tests. *R package version*, 14(1).
- 664 Link, W. A. and Nichols, J. D. (1994). On the importance of sampling variance to inves-  
665 tigations of temporal variation in animal population size. *Oikos*, 69(3):539 – 544.
- 666 Louthan, A. M., Keighron, M., Kiekebusch, E., Cayton, H., Terando, A., and Morris, W. F.  
667 (2022). Climate change weakens the impact of disturbance interval on the growth rate  
668 of natural populations of venus flytrap. *Ecological Monographs*, 92(4):e1528.
- 669 McGillivray, H. (1986). Skewness and asymmetry: measures and orderings. *Annals of*  
670 *Statistics*, 14:994–1011.
- 671 Metcalf, C. J. E., Ellner, S. P., Childs, D. Z., Salguero-Gómez, R., Merow, C., McMahon,  
672 S. M., Jongejans, E., and Rees, M. (2015). Statistical modelling of annual variation for  
673 inference on stochastic population dynamics using Integral Projection Models. *Methods*  
674 *in Ecology and Evolution*, 6:1007–1017.
- 675 Miller, T. E., Louda, S. M., Rose, K. A., and Eckberg, J. O. (2009). Impacts of insect  
676 herbivory on cactus population dynamics: experimental demography across an envi-  
677 ronmental gradient. *Ecological Monographs*, 79(1):155–172.
- 678 Ohm, J. R. and Miller, T. E. (2014). Balancing anti-herbivore benefits and anti-pollinator  
679 costs of defensive mutualists. *Ecology*, 95(10):2924–2935.
- 680 Ozgul, A., Childs, D. Z., Oli, M. K., Armitage, K. B., Blumstein, D. T., Olson, L. E., Tul-  
681 japurkar, S., and Coulson, T. (2010). Coupled dynamics of body mass and population  
682 growth in response to environmental change. *Nature*, 466(7305):482–485.
- 683 Peterson, M. L., Morris, W., Linares, C., and Doak, D. (2019). Improving structured  
684 population models with more realistic representations of non-normal growth. *Methods*  
685 *in Ecology and Evolution*, 10(9):1431–1444.
- 686 Plard, F., Schindler, S., Arlettaz, R., and Schaub, M. (2018). Sex-specific heterogene-  
687 ity in fixed morphological traits influences individual fitness in a monogamous bird  
688 population. *The American Naturalist*, 191(1):106–119.
- 689 Rees, M., Childs, D. Z., and Ellner, S. P. (2014). Building integral projection models: a  
690 user's guide. *Journal of Animal Ecology*, 83(3):528–545.

- 691 Salguero-Gómez, R. and Casper, B. B. (2010). Keeping plant shrinkage in the demo-  
692 graphic loop. *Journal of Ecology*, 98(2):312–323.
- 693 Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky,  
694 C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020).  
695 Robustness of linear mixed-effects models to violations of distributional assumptions.  
696 *Methods in ecology and evolution*, 11(9):1141–1152.
- 697 Schultz, E. L., Eckberg, J. O., Berg, S. S., Louda, S. M., and Miller, T. E. (2017). Native  
698 insect herbivory overwhelms context dependence to limit complex invasion dynamics  
699 of exotic weeds. *Ecology letters*, 20(11):1374–1384.
- 700 Snyder, R. E. and Ellner, S. P. (2018). Pluck or luck: Does trait variation or chance drive  
701 variation in lifetime reproductive success? *American Naturalist*, 191(4):E90 – E107.
- 702 Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location  
703 scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46.
- 704 Stubberud, M. W., Vindenes, Y., Vøllestad, L. A., Winfield, I. J., Stenseth, N. C., and Lan-  
705 gangen, Ø. (2019). Effects of size-and sex-selective harvesting: An integral projection  
706 model approach. *Ecology and Evolution*, 9(22):12556–12570.
- 707 Teller, B. J., Adler, P. B., Edwards, C. B., Hooker, G., and Ellner, S. P. (2016). Linking  
708 demography with drivers: climate and competition. *Methods in Ecology and Evolution*,  
709 7(2):171–183.
- 710 TJ, D. and AC, M. (2004). Inferential aspects of the skew exponential power distribution.  
711 *Journal of the American Statistical Association*, 99:439 – 450.
- 712 Tredennick, A. T., Teller, B. J., Adler, P. B., Hooker, G., and Ellner, S. P. (2018). Size-by-  
713 environment interactions: a neglected dimension of species' responses to environmen-  
714 tal variation. *Ecology Letters*, 21(12):1757–1770.
- 715 Wan, X., Wang, W., Liu, J., and Tong, T. (2014). Estimating the sample mean and stan-  
716 dard deviation from the sample size, median, range and/or interquartile range. *BMC*  
717 *medical research methodology*, 14:1–13.
- 718 Williams, J. L., Miller, T. E., and Ellner, S. P. (2012). Avoiding unintentional eviction from  
719 integral projection models. *Ecology*, 93(9):2008–2014.

<sub>720</sub> Wood, S. (2017). *Generalized Additive Models: An Introduction with R.* Chapman and  
<sub>721</sub> Hall/CRC, 2 edition.

<sub>722</sub> Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational  
723 Statistics & Data Analysis*, 79:277–291.

# Appendices

## S.1 The Jones-Pewsey distribution

Jones and Pewsey (2009) introduced a simple, tractable generalization of the Normal distribution with two additional parameters determining asymmetry (skewness), and tail weight (kurtosis) which can be either lighter or heavier than the Gaussian. It is defined as a transformation of a  $\text{Normal}(0,1)$  random variable using the hyperbolic sine function ( $\sinh$ ) and its inverse ( $\text{asinh}$ ), as follows. The distribution family's base probability density  $f_{\epsilon,\delta}$  is the probability density of the random variable  $X_{\epsilon,\delta}$  where

$$Z = \sinh(\delta \text{ asinh}(X_{\epsilon,\delta}) - \epsilon) \quad (\text{S.1})$$

and  $Z$  has a  $\text{Normal}(0,1)$  distribution. Equivalently,

$$X_{\epsilon,\delta} = \sinh\left(\frac{1}{\delta} \text{ asinh}(Z) + \frac{\epsilon}{\delta}\right). \quad (\text{S.2})$$

Parameters  $\delta = 1, \epsilon = 0$  give the  $\text{Normal}(0,1)$  distribution. Skewness has the sign of  $\epsilon$ , and  $\delta > 0$  controls tail weight, with heavier than Gaussian tails for  $\delta < 1$  and lighter than Gaussian tails for  $\delta > 1$ . A formula for the density  $f_{\epsilon,\delta}$  is given by Jones and Pewsey (2009, eqn. 2). The general four-parameter family with location parameter  $\mu$  and scale parameter  $\sigma$  is defined as the probability densities of  $\mu + \sigma X_{\epsilon,\delta}$ . We refer to this as the JP distribution family.

As is unfortunately the case for most four-parameter distributions  $\mu$  is not the mean,  $\sigma$  is not the standard deviation,  $\epsilon$  is not the skew and  $\delta$  is not the kurtosis. All else being equal, larger  $\mu$  gives a larger mean, larger  $\sigma$  gives a higher standard deviation, higher  $\epsilon$  gives higher asymmetry, and higher  $\delta$  gives heavier tail weight. But each moment is jointly determined by all four parameters.

The main advantage of the JP distribution is that the attainable combinations of skewness and kurtosis are very broad, compared to other four-parameter families, and come very close to the theoretical limits on kurtosis as a function of skewness (Jones and Pewsey, 2009, Fig. 2). Additionally, being a transformation of the Normal makes it very simple to generate random numbers from the distribution, and to compute probability density, cumulative distribution, and quantile functions. There are also simple analytic formulas for the first four moments (Jones and Pewsey, 2009, p. 764) which we use below

752 to define a centered and scaled version in which  $\mu$  and  $\sigma$  are the mean and standard  
753 deviation.

754 The definition (S.2) shows that the distribution depends on  $\epsilon$  only through the ratio  
755  $\epsilon/\delta$ . We have found that this property can be problematic for estimating distribution  
756 parameters. Even with good sized ( $n = 250$  or  $500$ ) data sets generated from the distri-  
757 bution with known parameters, both maximum likelihood and Bayesian estimation were  
758 unstable for some values of  $\epsilon$  and  $\delta$ , occasionally yielding estimates far from the truth.  
759 One cause was a ridge in the  $(\epsilon, \delta)$  likelihood surface with a constant of  $\epsilon/\delta$ . Another is  
760 that when  $\delta$  is large, changes in  $\epsilon$  have little effect.

761 To avoid that problems, we reparameterize the distribution as follows:

$$\text{X}_{\lambda, \tau} = \sinh(e^{-\tau} \operatorname{asinh}(Z) + \lambda). \quad (\text{S.3})$$

763 Thus, the two parameterizations are related by

$$\delta = e^\tau, \epsilon = \delta\lambda = e^\tau\lambda. \quad (\text{S.4})$$

765 The definition of  $\tau$  allows it to take any real value, with negative values giving thinner  
766 than Gaussian tails and positive values giving fatter than Gaussian tails.  $\lambda$  also can take  
767 any real value, and the distribution's skew has the same sign as  $\lambda$ . Because the  $\sinh$   
768 function is nonlinear, it is still the case that the skew depends on  $\tau$  as well as  $\lambda$ , but the  
769 "crosstalk" between the kurtosis and skew parameters is weaker. As a result, we found  
770 that maximum likelihood estimation of parameter values was generally more reliable if  
771 the distribution is parameterized in terms of  $\tau$  and  $\lambda$ .

## 772 S.2 Estimating mixed-effects models using shrinkage

773 Ecologists often fit demographic and other statistical models that include random effects  
774 terms to quantify variation among years, spatial locations, individuals, etc. Random  
775 effects are a natural choice when interest centers on the magnitude of variation (e.g., how  
776 much does mortality vary among years?) rather than individual values (e.g., mortality  
777 in 2013). They also allow each estimate to "borrows strength" from others, so that (for  
778 example) the estimate from a year with small sample size (and thus large sampling  
779 variability) is shifted towards the center of the overall distribution.

780 Specialized software is often used to fit such models, such as the **nlme**, **lme4**, **mgee**  
781 and **gamm4** libraries in R, but these only allow a small subset of the distribution families

782 we want to consider for modeling growth increments (the **gamlss** package allows many  
783 distribution families, but in our experience, even when random effects are simple in  
784 structure the fitting algorithms often fail to converge or fail to find the global optimum).

785 One way past this limitation is Bayesian estimation, using STAN with user-written  
786 (or borrowed) code for the chosen growth distribution (see section XX for an example).  
787 In this appendix we describe another option, introduced by Link and Nichols (1994)  
788 and Gould and Nichols (1998): fitting a fixed-effects model by Maximum Likelihood,  
789 followed by shrinkage of coefficient estimates. None of the ideas here are original. The  
790 material overlaps Appendix S1 of Metcalf et al. (2015), but for completeness we make  
791 it self-contained. Appendix D of Cooch and White (2020) (written by K.D. Burnham)  
792 provides more details and examples in the context of capture-recapture analysis.

793 Here we explain shrinkage using a simple model based on our analysis of *Pseu-*  
794 *doroegneria spicata*. That model includes random effects for between-year variation in  
795 the slope and intercept of future size (log area) as a function of initial size. To keep  
796 the example simple, we assume that initial size and year are the only covariates, and  
797 we assume that growth increments follow a skew-Normal distribution with noncon-  
798 stant variance and constant skew parameter. Code for this example is in the script  
799 `SimpleShrinkageExample.R`. The first part of the script generates an artificial data set  
800 by fitting the model to a subset of the growth data (20th century Control plots), and  
801 randomly generating new “size next year” values for each individual in the actual data  
802 set. The second part contains the “data” analysis.

803 As in our *P. spicata* analysis, we assumed that that the skew and kurtosis parameters  
804 were functions of the location parameter; this dominated ( $\Delta AIC \approx 30$ ) the alternate  
805 model with skew and kurtosis depending on initial size. The analogous Gaussian model,  
806 with constant variance, could be fitted as follows using `lmer`:

807 `lmer(new.size ~ init.size + (init.size|year), data=growthData, REML=TRUE);`  
808 where `growthData` is a data frame holding the data with `year` as an unordered factor.  
809 For our skew-Normal model, we instead use maximum likelihood with all between-year  
810 variation included as fixed effects. The appropriate design matrix is easily constructed  
811 using the `model.matrix` function:

812 `U = model.matrix(~ year + init.size:year - 1, data=growthData)`

813 If there are  $T$  years, the matrix `U` specified in this way has  $2T$  columns corresponding to  
814  $n$  annual intercepts and  $T$  annual slopes.

Using this design matrix, we can readily write a log likelihood function for use with the **maxLik** package, with a log link function for the variance because it is necessarily positive:

```
815 LogLik=function(pars,new.size,U){  
816     pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];  
817     mu = U%*%pars1;  
818     sigma = exp(pars2[1]+pars2[2]*mu);  
819     dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)  
820 }  
821  
822 }
```

Parameters and their standard errors can then be estimated with **maxLik**, starting from a random guess:

```
823 start=c(runif(ncol(U)), rep(0,3))  
824 out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,  
825     method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);  
826 coefs = out$estimate; # parameters  
827 V = vcov(out); SEs = sqrt(diag(V)); # standard errors
```

In real life we would repeat the optimization several times with several different starting values, to be confident that the optimal parameter values had been found.

Focus now on the year-specific intercept parameters  $\hat{a}_t, t = 1, 2, \dots, T$ . We can view the year-specific estimates  $\hat{a}_t$  as consisting of unobserved true values  $a_t$  plus sampling error:

$$\hat{a}_t = a_t + \varepsilon_t \quad (\text{S.5})$$

Because of the sampling errors, the sample variance of the estimates  $\hat{a}_t$  is an upward-biased estimate of the true across-year variance in the parameter. That is undesirable if the model will be used to project how temporal variability affects population dynamics. However, maximum likelihood estimation gives us an approximate variance-covariance matrix  $\hat{V}$  of the sampling errors,  $V$  in the code above. With that information, we can estimate the parameters of a random effects model for the intercept parameters, and thereby improve the year-specific estimates and the estimate of the across-year variance.

The model is as follows. We make the standard mixed-models assumptions that the  $a_t$  are drawn independently from some fixed distribution with unknown variance  $\sigma^2$ . We also assume that the estimates  $\hat{a}_t$  are unbiased, that is

$$\mathbb{E}(\varepsilon_t | a_t) = 0. \quad (\text{S.6})$$

848 These are optimistic assumptions, but not excessively optimistic. Some degree of tem-  
 849 poral correlation will often be present, and as we explain at the end, it is theoretically  
 850 possible to account for it. Maximum likelihood parameter estimates are not unbiased,  
 851 but if the assumptions of maximum likelihood are satisfied the bias is asymptotically  
 852 negligible compared to the standard error (the bias scales as the inverse of sample size,  
 853 the standard error as the square root of the inverse of sample size).

854 Let  $S^2$  denote the sample variance of the estimates  $\hat{a}_t$ . It can then be shown that

$$855 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

856 This is eqn. (1) in Gould and Nichols (1998) in our notation, without the term that results  
 857 from temporal autocorrelation.

858 The terms besides  $\sigma^2$  on the right-hand are the expected impact of sampling error  
 859 on the across-year variance of the parameter estimates; their presence makes  $S^2$  a biased  
 860 estimate of  $\sigma^2$ . However, all of those terms correspond to entries in the variance-  
 861 covariance matrix  $V$ . We can therefore use our estimated variance-covariance matrix  $\hat{V}$   
 862 to remove the bias due to sampling variability:

$$863 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

864  $\hat{\sigma}^2$  estimates the variance of the distribution from which the  $a_t$  are assumed to be drawn.

865 Using that estimate, we can adjust the year-specific estimates to reduce the ex-  
 866 pected impact of sampling error. Depending on your purposes, there are two possible  
 867 adjustments. The first option is the one used in the popular capture-recapture analysis  
 868 software Mark Cooch and White (2020),

$$869 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

870 The name “shrinkage” comes from the fact that each estimate is adjusted towards the  
 871 overall mean, with larger adjustments of values that have higher estimated sampling  
 872 error variance,  $\hat{V}_{t,t}$ . This shrinkage estimate has the property that the expected sample  
 873 variance of the adjusted estimates  $\tilde{a}_t$  is very close to  $\hat{\sigma}^2$ , so the  $\tilde{a}_t$  approximate the actual  
 874 amount of parameter variation.

875        The second is to replace  $\hat{a}_t$  by the least-squares estimate of  $a_t$  under the additional  
 876        assumption that the  $a_t$  are drawn from a Gaussian distribution; this is given by

$$877 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.10})$$

878        This option is theoretically preferable if the Gaussian assumption is reasonable, and you  
 879        are more interested in year-specific values rather than across-year variance. However,  
 880        Metcalf et al. (2015) found that even (S.9), which does less shrinkage, resulted in a small  
 881        downward bias in the temporal variance of population growth rates. This argues for  
 882        always using the first option, and we do the same here.

883        We differ from MARK, however, in using (S.8) rather than an iterative method  
 884        that takes (S.8) as its starting estimate and refines the estimate by using weighted least  
 885        squares based on the current estimate. Metcalf et al. (2015) found, in simulation studies,  
 886        that the iterative method was either slightly beneficial or wildly inaccurate. We therefore  
 887        advise against it.

888        Finally, as mentioned above, the estimate of  $\sigma^2$  can account for temporal autocor-  
 889        relation in the  $a_t$ . When present, those correlations add a term to eqn. (S.7) (see eqn.  
 890        (1) in Gould and Nichols (1998)), which can be estimated from the sample autocorre-  
 891        lation of the  $\hat{a}_t$ . We do not recommend doing this (and therefore omit the formulas)  
 892        because the autocorrelations can only be reliably estimated if they fall to nearly zero  
 893        within lag  $m \ll T$ , in which case the autocorrelation term is small (specifically,  $O(m/T)$ ).  
 894        Otherwise, the random error from using poorly estimated autocorrelations is likely to  
 895        outweigh the small bias from omitting that term.

896        The take-home message is that estimating random effects from the regression coef-  
 897        ficients is very simple:

```
898 # Variance-covariance matrices for intercepts and slopes
899 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
900 # Extract year-specific intercepts, center them to zero
901 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
902
903 # Estimate sigma^2
904 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
905           (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
906
907 # Shrink deviations from the mean
```

```

908 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
909
910 # Do it all again for the slopes
911 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
912 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
913   (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
914 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

915 The figure below shows the results for one artificial PSSP “data” set, having  $T = 22$   
916 years and growth measurements on about 175 individuals/year on average. The true  
917 random year effects (the ones used to generate the data) are recovered with good accu-  
918 racy and no bias. In particular there is no sign of extreme values being pulled in too  
919 far towards the mean, which would cause an S-shaped graph of estimated versus true  
920 values.

