

# 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. Integral Projection Models (IPMs) are widely used for studying the dynamics of  
<sup>3</sup> continuously size-structure populations. IPMs require a growth sub-model that  
<sup>4</sup> describes the probability of future size conditional on current size. Over the past  
<sup>5</sup> two decades, most IPM studies have assumed that this probability is normally-  
<sup>6</sup> distributed, despite repeated calls for non-Gaussian approaches that accommodate  
<sup>7</sup> skewness and kurtosis known to occur in size transition data.
- <sup>8</sup> 2. We provide a general workflow for modeling size transitions that accommodates  
<sup>9</sup> non-Gaussian growth patterns while retaining the desirable features (ecologically  
<sup>10</sup> important covariates and random effects) that Gaussian approaches typically pro-  
<sup>11</sup> vide. Our approach emphasizes visual diagnostics of residuals from pilot Gaussian  
<sup>12</sup> models and quantile-based metrics of skewness and kurtosis that vet the fit of the  
<sup>13</sup> Gaussian distribution and guide the selection of an alternative, if necessary. We  
<sup>14</sup> illustrate our methods by reanalyzing size transition data from our published IPM  
<sup>15</sup> studies, targeting a diversity of demographic quantities including population growth  
<sup>16</sup> rate, invasion wave velocity, and evolutionarily stable life history strategies.
- <sup>17</sup> 3. Across one coral and three plant case studies, skewness and excess kurtosis were  
<sup>18</sup> common features of size transition data and non-Gaussian growth models consis-  
<sup>19</sup> tently generated simulated data that were more consistent with the real data than  
<sup>20</sup> pilot Gaussian models. However, in these case studies, the effects of “improved”  
<sup>21</sup> growth modeling on IPM results were generally modest, and differed in direction or  
<sup>22</sup> magnitude between different outputs from the same model.
- <sup>23</sup> 4. Using tools that were not available when IPMs were first developed, it is now possi-  
<sup>24</sup> ble to fit non-Gaussian models to size transition data; our worked examples demon-  
<sup>25</sup> strate how, including open-access data and computing scripts. Doing so, as guided  
<sup>26</sup> by careful interrogation of the data, will result in a model that better represents the  
<sup>27</sup> population for which it is intended.

<sup>28</sup> **Keywords**

## 29 Introduction

30 Structured demographic models – matrix and integral projection models (MPMs and  
31 IPMs) – are powerful tools for data-driven modeling of population dynamics and viability  
32 that are widely used in basic and applied settings. In contrast to MPMs for populations with discrete structure (life stage, age class, etc.), IPMs (Easterling et al., 2000) readily  
33 accommodate populations structured by continuous state variables, most commonly size.  
34 A related innovation of the IPM framework is its emphasis on regression-based  
35 modeling for parameter estimation, which carries important advantages for making the  
36 most of hard-won data (Ellner et al., 2022).

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

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

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

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

The observation that the normal distribution may poorly describe size transitions in real organisms has been made before, and several studies have emphasized that alternative distributions should be explored (Easterling et al., 2000; Peterson et al., 2019; Rees et al., 2014; Williams et al., 2012). Yet, default use of Gaussian growth distributions (often with non-constant variance) remains the standard practice. The general state-of-the-art in the literature appears to remain where it was 20 or so years ago, using the default model without pausing to examine critically whether or not it actually provides a good description of the data. We are guilty of this, ourselves.

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

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

Our goal here is to present and illustrate a general 'recipe' that moves growth modeling past the standards set over 20 years ago. Like any recipe, users may need to make substitutions or add ingredients to suit their situation. Our approach emphasizes graphical diagnostics for developing and evaluating growth models, rather than a process centered on statistical model selection. Through a set of empirical case studies we demonstrate how a simple workflow, using tools that were nonexistent or not readily available when IPMs first came into use, makes it straightforward and relatively easy to

100 identify when the default model is a poor fit to the data, and to then choose and fit a  
101 substantially better growth model that is no harder to use in practice. We illustrate our  
102 approach by revisiting four of our own, mostly published IPM analyses that assumed  
103 Gaussian growth. In each case, the Gaussian assumption does not stand up to close  
104 scrutiny. We illustrate how we could have done better, and the consequences of “doing  
105 better” for our ecological inferences. All of our analyses may be reproduced from code  
106 and data that are publicly available (see Data accessibility statement).

## 107 A general workflow for better growth modeling

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

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

111 This step is familiar to most IPM users, as it is the start and end of the traditional  
112 workflow. A well-fitted Gaussian model accurately describes the mean and variance  
113 of future size conditional on current size and possibly on other measured covariates  
114 or random effects. This step may include model selection to identify which treat-  
115 ment effects or environmental drivers affect the mean and/or variance of future size.  
116 Non-constant variance is often fitted in a two-stage process, first fitting mean growth  
117 assuming constant variance, then doing a regression relating the squared residuals  
118 from the initial fit to the fitted mean. It is sometimes better to fit size-dependence  
119 in the mean and variance simultaneously, as can be done with the R packages **mgcv**  
120 and **nmle**, because incorrectly assuming constant variance can affect the outcome of  
121 model selection for the mean. One-step fitting is straightforward for simple models  
122 in which initial size is the only factor that can influence growth variance. However,  
123 the two-step process fitting residuals to the fitted value (expected future size) may  
124 be convenient when there are multiple fixed and random effects, all of which may  
125 contribute to non-constant variance, since the expected value implicitly accounts for  
126 all of them. We illustrate both one-step and two-step approaches in the examples  
127 below.

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

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

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

140     There are many ways that growth data may deviate from Gaussian, and the na-  
141     ture of those deviations can guide the search for a better distribution. Frequentist  
142     tests such as the D'Agostino test of skewness (D'Agostino, 1970) and the Anscombe-  
143     Glynn test of kurtosis (Anscombe and Glynn, 1983) could be used to diagnose  
144     whether the aggregate distribution of standardized residuals deviates from normal-  
145     ity (R package **moments** (Komsta and Novomestky, 2015)). However, the aggregate  
146     distribution of standardized residuals may be misleading if properties such as skew  
147     and kurtosis vary with size. For example, a change in the direction of skewness from  
148     small to large sizes would require a distribution flexible enough to accommodate  
149     both positive and negative skew, such as the skewed normal or Johnson  $S_U$  distri-  
150     butions. Alternatively, growth data may lack skew but may exhibit leptokurtosis (in  
151     which case the  $t$  distribution may be a good choice) or may shift from platykurtosis  
152     to leptokurtosis depending on initial size (in which case the power exponential  
153     distribution may be a good choice). It is therefore essential to visualize trends in dis-  
154     tribution properties with respect to size, either initial size (for simple models with  
155     only size-dependence) or expected future size (for models with multiple fixed ef-  
156     fects). In the case studies below, we rely on quantile regression of the standardized  
157     residuals to visualize skew and kurtosis as continuous functions of size or expected  
158     value. Fig. 1 includes guidance on how the skew and kurtosis properties of the stan-  
159     dardized residuals suggest options for an appropriate growth distribution. In our  
160     case studies we take advantage of the many distributions provided in the **gamlss** R  
161     package (Stasinopoulos et al., 2007), but any other distributions with the necessary  
162     properties can be used.

- 163     3. Refit the growth model using the chosen distribution.

164     In models with multiple covariates and/or random effects, each potentially affecting  
165     several distribution parameters (location, scale, skew, kurtosis) in different ways,  
166     “refit the model” could entail a massive model selection process to identify the  
167     “right” or “best” non-Gaussian model. And with so many options, model uncer-

tainty may be overwhelming and over-fitting becomes a significant risk even if precautions against it are taken. We therefore argue for adopting the more modest goal of remedying the apparent defects in the Gaussian model. Conveniently, as we demonstrate below, the functional forms for the mean and standard deviation (or location and scale parameters) could be carried over from the pilot Gaussian model into a non-Gaussian distribution, leaving skew and kurtosis as the targets for improvement. This step exploits the fact that parameter estimation from a Gaussian model is generally robust to deviations from normality (Schielzeth et al., 2020), meaning that the mean of the Gaussian model is probably a good proxy for the mean of the non-Gaussian model (and in case it is not, the next step in the workflow would catch that). The functional forms for skew and kurtosis of the non-Gaussian model can be guided by the qualitative features of the graphical diagnostics (e.g., skewness switches from positive to negative with size).

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

## How should skewness and kurtosis be measured?

“Improvement” of a Gaussian model will always involve scrutiny of skewness and kurtosis, so measurement of these properties warrants some attention. The standard measures of skewness and kurtosis (tail thickness) are based on the third and fourth central moments, respectively, of the distribution:

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

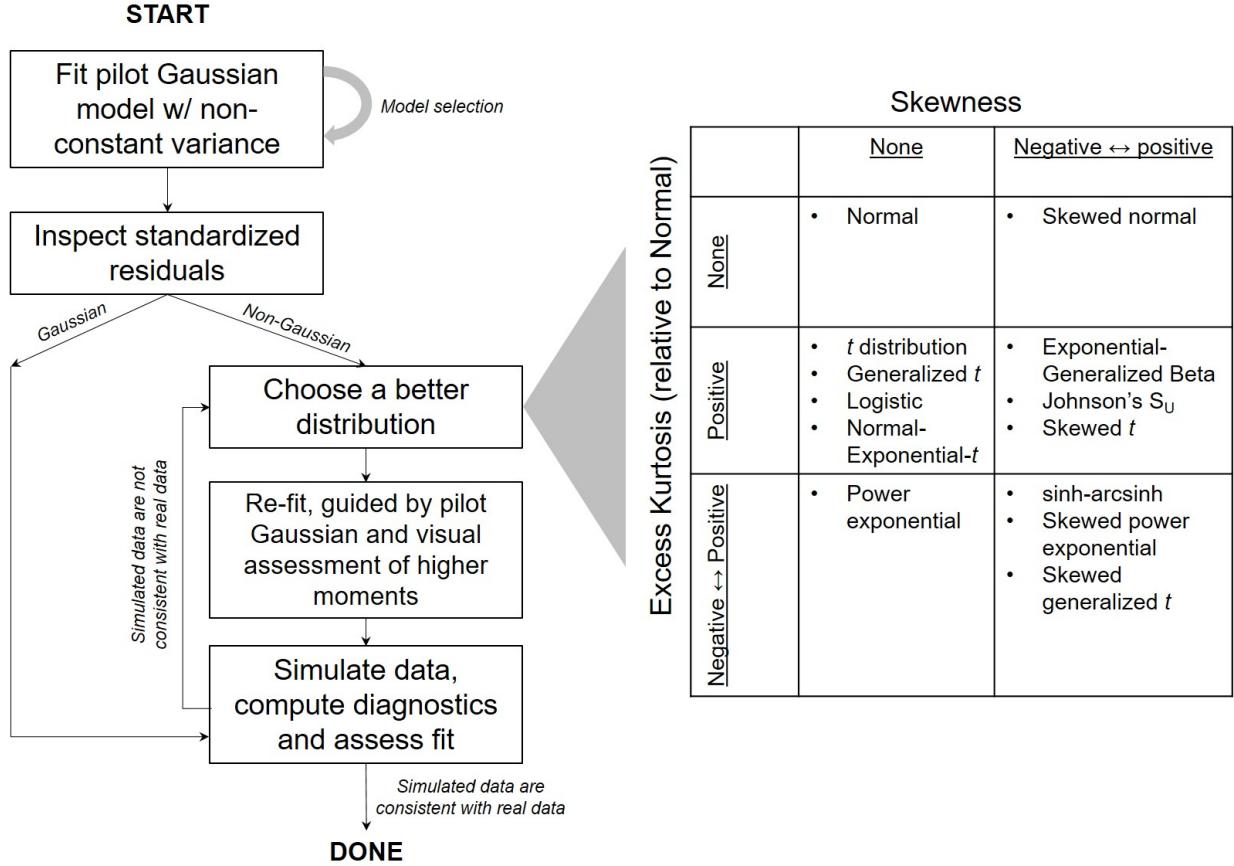


Figure 1: General workflow of recommendations for IPM growth modeling (left) and guide to common non-Gaussian distributions of size  $x$  for  $x \in \mathbb{R}$  that can accommodate different combinations of skewness and kurtosis (right). All of these distributions (often including multiple versions 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).

199 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  
 200 variance (second central moment). A Gaussian distribution has zero skewness and zero  
 201 excess kurtosis.

202 The standard measures are easy to calculate but their use for choosing and evaluating  
 203 growth models is hindered by their poor sampling properties. Because empirical  
 204 estimates involve high powers of data values, it only takes a few outliers to produce  
 205 a very inaccurate estimate. Figure 2 shows a simulated example, where the underlying  
 206 “data” are a sample of size 200 from a  $t$  distribution with 8 degrees of freedom; the true  
 207 skew is 0, and the true excess kurtosis is 1.5. The distance between the largest and small-  
 208 est estimates (indicated by the dotted red vertical lines), relative to the distance between



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`

209 the 5th and 95th percentiles, shows the broad extent of extreme values that can occur  
 210 even with a good size sample, especially for kurtosis.

211 We therefore use “nonparametric” (NP) measures of skew and kurtosis that are  
 212 based on quantiles and thus less sensitive to a few extreme data values. Let  $q_\alpha$  denote  
 213 the  $\alpha$  quantile of a distribution or sample (e.g.,  $q_{0.05}$  is the 5th percentile). For any  
 214  $0 < \alpha < 0.5$ , a quantile-based measure of skewness is given by (McGillivray, 1986)

$$215 \quad \text{NP Skewness} = \frac{q_\alpha + q_{1-\alpha} - 2q_{0.5}}{q_{1-\alpha} - q_\alpha}. \quad (2)$$

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

median. The lower tail size is  $\tau_L = q_{0.5} - q_\alpha$ . The definition above is equivalent to

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

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

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

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

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

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

Quantile-based estimation of skewness and kurtosis carries the added value that quantile regression methods may be used to derive these properties of size transitions as continuous functions of initial size or expected future size. In the examples below, we use the **qgam** package to fit smooth additive quantile regression models, which have the flexibility to accommodate non-linear size-dependence in skewness and kurtosis. One risk of a gam-based approach is that fitted quantiles may be too “wiggly” without 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

251 to visualize these as functions of size. Specifically, following Wan et al. (2014),

252

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

253 and

254

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

255 

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

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

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

```
278 # XH is a data frame holding the data
279 # logarea.t0, .t1 denote initial and final values of log-transformed area
280 fitGAU <- gam(list(logarea.t1~ s(logarea.t0), ~ s(logarea.t0)),
281             data=XH, gamma=1.4, family=gaulss())
282 fitted_all = predict(fitGAU,type="response");
```

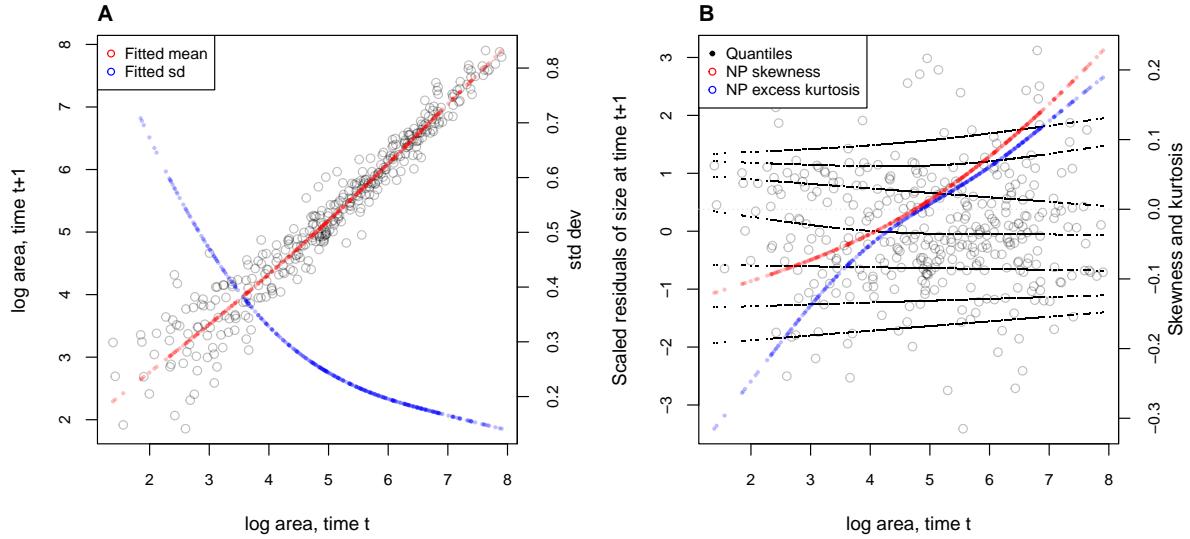


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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.

```

283 fitted_sd = 1/fitted_all[,2];
284 scaledResids = residuals(fitGAU,type='response')/fitted_sd;

```

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

289 There are no blatant signs of trouble in the pilot Gaussian model, but quantile re-  
290 gressions on the scaled residuals, and the NP Skewness and Kurtosis metrics derived  
291 from them (Eq. 3 and 4), suggest deviations from normality (Fig. 3B). Specifically,  
292 skewness switches from negative to positive across the size distribution, with smaller  
293 corals more likely to shrink than grow and larger corals more likely to grow than shrink.  
294 Kurtosis also changes direction over the size distribution, with smaller initial sizes hav-  
295 ing thinner tails and larger initial sizes having fatter tails than Gaussian. The fitted  
296 nonparametric moments suggest that the upper and lower tails of size transition proba-  
297 bilities may differ by up to 20%, and the weight of the tails may be 20% greater or less  
298 than Gaussian, depending on initial size – not overwhelming deficiencies, but not trivial  
299 either. Are these deviations from normality severe enough to warrant a second, non-

300 Gaussian iteration of growth modeling? This question may be answered by simulating  
301 data from the Gaussian model and examining whether key properties of the simulated  
302 data are consistent with those of the real data – this is the ultimate litmus test for a  
303 growth model’s adequacy and should be a standard element of IPM construction, in our  
304 opinion. If the simulated data are not consistent with the real data, it is time to choose  
305 a better distribution (Fig. 1). In this case, the negative skew at small sizes and excess  
306 kurtosis observed at large sizes are more extreme than what occurs across 100 random  
307 iterations of data simulation (Fig. 4), suggesting that, for at least some parts of the size  
308 distribution, a non-Gaussian model would better capture size transitions.

309 We sought a distribution that could accommodate the properties of the scaled resid-  
310 uals, specifically changes in the sign of skewness and excess kurtosis across initial sizes.  
311 We chose the sinh-arcsinh (SHASH) distribution, a four-parameter distribution that, con-  
312 veniently, is included in **mgcv**’s **gam()** function:

```
313 fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0,k=4), # <- location  
314 ~ s(logarea.t0,k=4), # <- log-scale  
315 ~ s(logarea.t0,k=4), # <- skewness  
316 ~ s(logarea.t0,k=4)), # <- log-kurtosis  
317 data = XH, family = shash, optimizer = "efs")
```

318 Data simulated from this model are more consistent with the real data than the Gaussian  
319 model: many of the 100 simulated SHASH data sets exhibited negative skew at small  
320 sizes and positive excess kurtosis at large sizes that were as strong or stronger than  
321 observed in the real data (Fig. 4). If one cared to quantify the difference between models,  
322 the SHASH is clearly favored by AIC despite having twice as many parameters as the  
323 Gaussian ( $\Delta AIC = 7.04$ ).

324 What, then, have we gained by fitting a better growth model? Fig. 5A compares  
325 the predicted distributions of subsequent size in the fitted model and Gaussian pilot  
326 models, for the median size of a new recruit (leftmost pair of curves), the median initial  
327 size (central curves), and the 95th percentile of initial size in the data (rightmost curves).  
328 The differences are small, and most pronounced for the smallest size, where recruits  
329 are predicted to grow slightly larger under the SHASH model than the Gaussian model.  
330 The direction of this difference was surprising, since the SHASH accommodates negative  
331 skew at small sizes in the data. However, in modeling skew appropriately, the SHASH  
332 model also gives a better prediction for mean growth at small sizes than the Gaussian

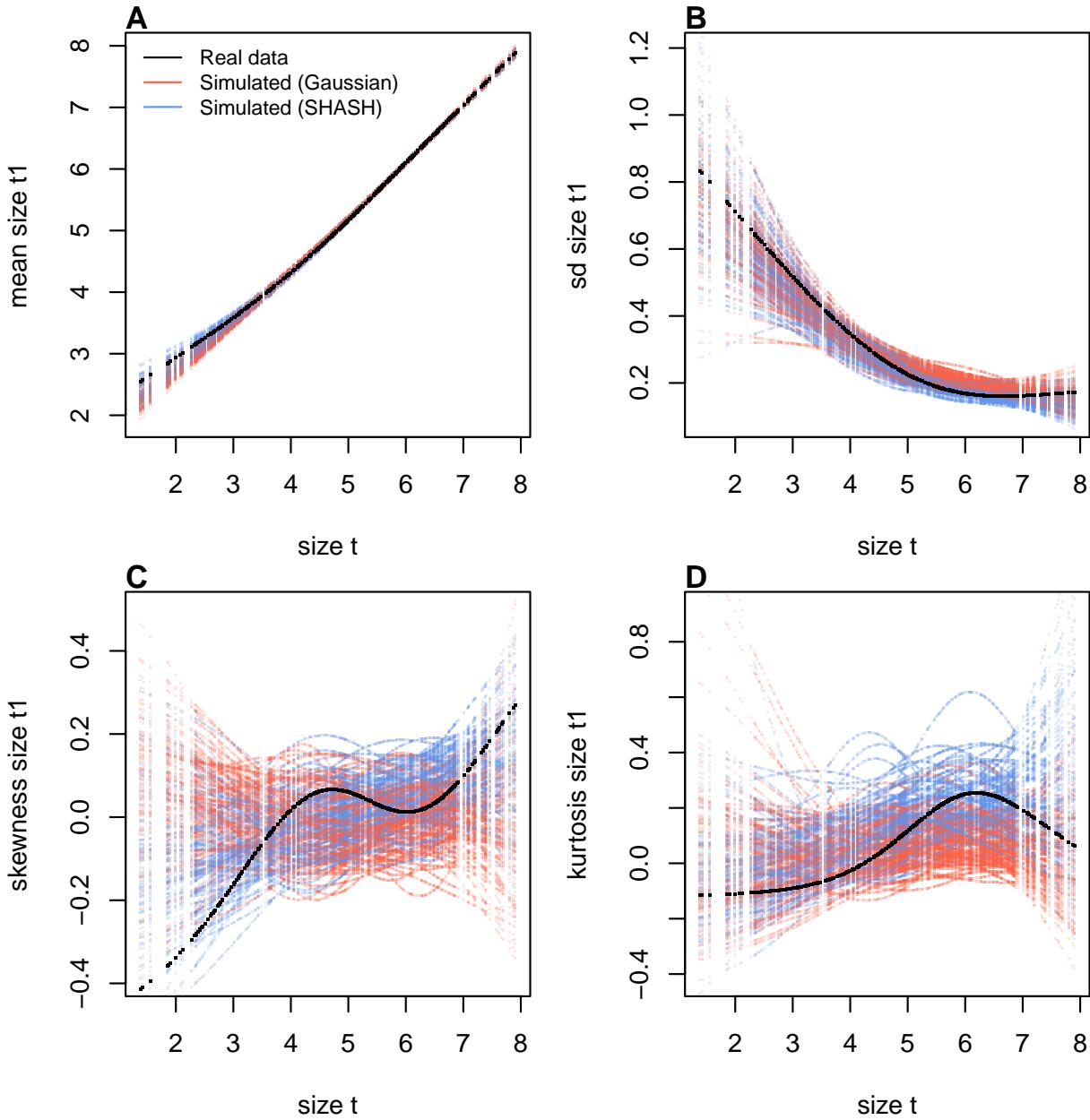


Figure 4: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script AkumalCorals\_qgam.R.

model, whose mean is biased downward by negative skew (Fig. 4A)<sup>1</sup>. Something similar happens in the standard deviation at large sizes (log size 5–7), where excess kurtosis in the data biased the SD upward (Fig. 4B). Fig. 5B shows the predicted steady-state size distributions resulting from a constant unit input of recruits. Again, the differences are

<sup>1</sup>...Contradicting the earlier assertion that parameter estimates from Gaussian models are robust to deviations from normality!

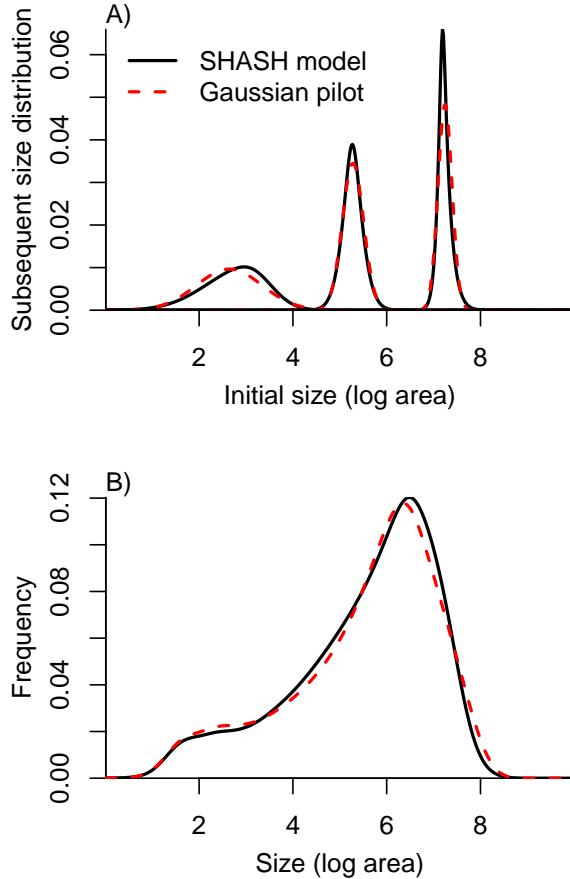


Figure 5: Comparisons between the fitted SEP1 growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year  $t + 1$  for three different values of size in year  $t$ . The leftmost pair of curves are 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.

337 very subtle. Finally, the Gaussian and SHASH growth models predict very similar mean  
 338 life span (17.7 and 17.9 years, respectively). From these outputs, there is little evidence  
 339 that improved modeling of coral growth meaningfully improved biological inferences  
 340 from the IPM; one could argue that it was not worth the trouble.

341 In this case study we used `gam` to fit both the Gaussian and SHASH models because  
342 that obviated model selection on functions for mean, variance, and higher moments.  
343 However, `gam` should be used with caution. Nonparametric regression models notori-  
344 ously “wag their tails” because the ends of the fitted curve can be pulled close to the  
345 outermost data points. This is especially problematic for growth modeling, because data  
346 are typically sparse near the bounds of the size distribution. To minimize the risk of  
347 overfitting we specified the number of “knots” (`k=4`) and used `gamma=1.4` to overweight  
348 model degrees of freedom, as suggested by Gu (2013, sec. 3.2). But it is always impor-  
349 tant to plot the fitted splines and make sure they do not wag unrealistically. If they do,  
350 parametric regression may be a better choice.

## 351 2 Case study: tree cholla cactus, *Cylindriopuntia imbricata*

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

372 The data for this new analysis include 4844 size transition observations from 929 in-  
373 dividuals spanning 13 transition years (2004–2018) and 11 spatial replicates (three spatial  
374 blocks in years 2004–2008 and eight 30m-by-30m plots in years 2009–2018). The data are  
375 provided in Miller (2020). Following previous studies, we quantified size as the natural  
376 logarithm of plant volume ( $cm^3$ ), derived from height and width measurements.

377 We begin the growth modeling workflow, as above, with a generalized additive  
378 model with the mean and standard deviation of size in year  $t + 1$  modeled as function  
379 of size in year  $t$ , with random intercepts for year and plot and assuming normally dis-  
380 tributed residuals (`family=gaulss()`). The standardized residuals, accounting for size-  
381 dependent residual variance (Fig. 6A), show clear signals of negative skew and positive  
382 excess kurtosis across most of the size distribution but strongest in the middle of the size  
383 distribution (Fig. 6B).

384 To better capture size transitions, we need a distribution with negative skew and  
385 positive excess kurtosis, but both of which may be negligible at some sizes. We first tried

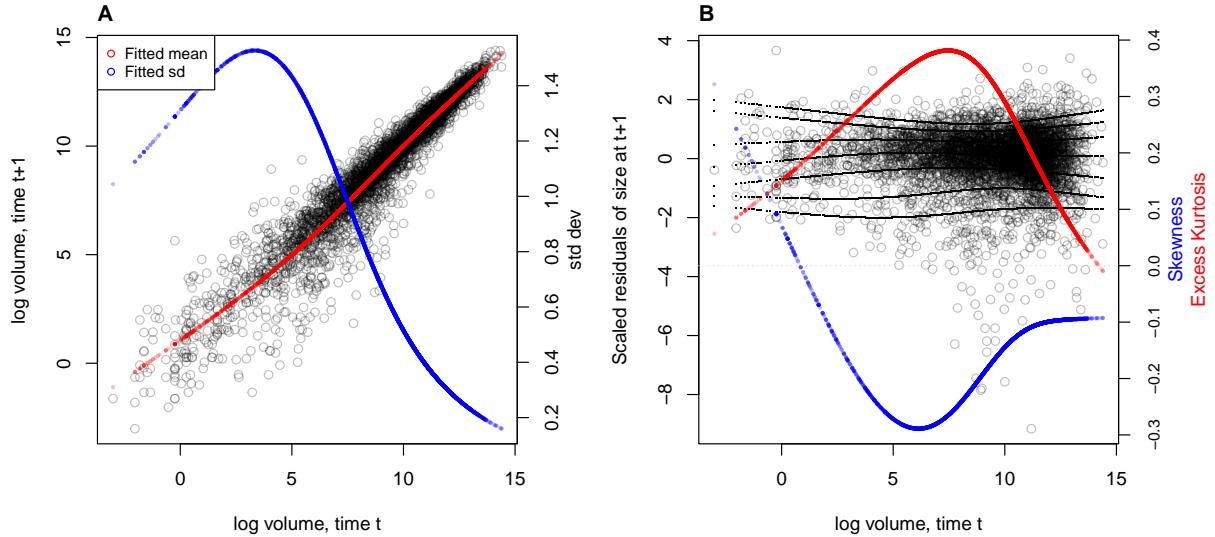


Figure 6: **A**, Size transition data for tree cholla cacti, *Cylindropuntia imbricata*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (blue) derived from them. Black lines in **B** show the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles. Figure made by script `cactus_growth_modeling_qgam.R`.

386 Johnson's  $S_U$  and then the skewed  $t$  distributions, both of which are limited to positive  
 387 excess kurtosis. Both distributions provided some improvement over the Gaussian, but  
 388 were not happy with the fit of either. Iterating through the workflow (Fig. 1), we ar-  
 389 rived, again, at the SHASH distribution, which is more flexible than either the JSU or  
 390 skewed  $t$ , capable of capturing a greater range of kurtosis for a given amount of skew,  
 391 and vice versa (Steve's NPSkewKurtosisRanges.pdf). Furthermore, fitting the SHASH  
 392 as a generalized additive model with `mgcv` allowed for flexible, non-monotonic size-  
 393 dependence in skewness and kurtosis without the need for model selection on specific  
 394 size-dependent functions; through iterations of trial and error, we found this flexibility  
 395 was necessary to generate simulated data that compared favorably to the real data. The  
 396 other distributions that we tried are not available as `mgcv` families, so we fit these with  
 397 custom maximum likelihood functions, an approach we illustrate in the next case study.  
 398 The final growth model was similar to the SHASH gam in the coral case study, but  
 399 with random intercepts for the location parameter, representing spatial and temporal  
 400 heterogeneity:

```
401 fit_shash <- gam(list(logvol_t1 ~ s(logvol_t,k=4) +  

402 s(plot,bs="re") + s(year_t,bs="re")), # <- model for locat
```

```

403 ~ s(logvol_t,k=4), # <- model for log-scale
404 ~ s(logvol_t,k=4), # <- model for skewness
405 ~ s(logvol_t,k=4)), # <- model for log-kurtosis
406 data = CYIM_grow,
407 family = shash,
408 optimizer = "efs")

```

409 The final SHASH model provided good correspondence between simulated and  
410 real data, and provided more compelling improvement over the Gaussian model than  
411 we saw in the coral case study (Fig. 7). The SHASH model over-estimated negative  
412 skew at some sizes relative to the signal of skewness in the data (Fig. 7C), but the nature  
413 of size-dependent skew in the data is not very biologically plausible and may instead  
414 be driven by the tail-wagging tendency of gams. As in the coral case study, we see  
415 that correctly modeling skewness and kurtosis improved estimation of the mean and  
416 standard deviation (Fig. 7A,B), yielding a growth model that is clearly truer to the data  
417 than the pilot Gaussian fit.

418 We explored how improved growth modeling influenced IPM results, leveraging  
419 the plot and year structure of the study design to quantify spatial and temporal vari-  
420 ance in fitness. We used the fitted random effects from the vital rate models to estimate  
421 the asymptotic growth rate for each year ( $\lambda_t$ ), centered on the average plot, and for  
422 each plot ( $\lambda_p$ ), centered on the average year. This allowed us to quantify demographic  
423 variance associated with temporal and spatial heterogeneity. We found that the Gaus-  
424 sian growth model tended to over-estimate  $\lambda_t$ , particularly in the harshest years (Fig.  
425 8A), and thus under-estimated temporal variance in fitness ( $Var(\lambda_{t(Gaussian)}) = 0.0018$ ,  
426  $Var(\lambda_{t(SHASH)}) = 0.0023$ ). The opposite was true for plot-to-plot variation (Fig. 8B),  
427 where the Gaussian model under-estimated  $\lambda_p$  and over-estimated spatial variance in  
428 fitness ( $Var(\lambda_{p(Gaussian)}) = 0.00015$ ,  $Var(\lambda_{p(SHASH)}) = 0.000088$ ). Across both growth  
429 models, fluctuations in fitness were stronger through time than across space. The  
430 difference in temporal variance would suggest that Gaussian growth modeling would  
431 lead to over-estimation of the stochastic growth rate  $\lambda_S$ , since temporal variance has  
432 a negative influence on  $\lambda_S$ . However, this was not the case: stochastic IPMs based  
433 on Gaussian and SHASH growth models had nearly identical stochastic growth rates  
434 ( $\lambda_S(Gaussian) = 0.9906$ ,  $\lambda_S(Gaussian) = 0.9909$ ). This is likely because temporal fluctu-  
435 ations in vital rates, which is where the SHASH growth model would make a difference,  
436 have a weaker influence on  $\lambda_S$  than the temporal fluctuations in size structure that they  
437 generate (Compagnoni et al., 2016; Ellis and Crone, 2013). Thus, depending on the target

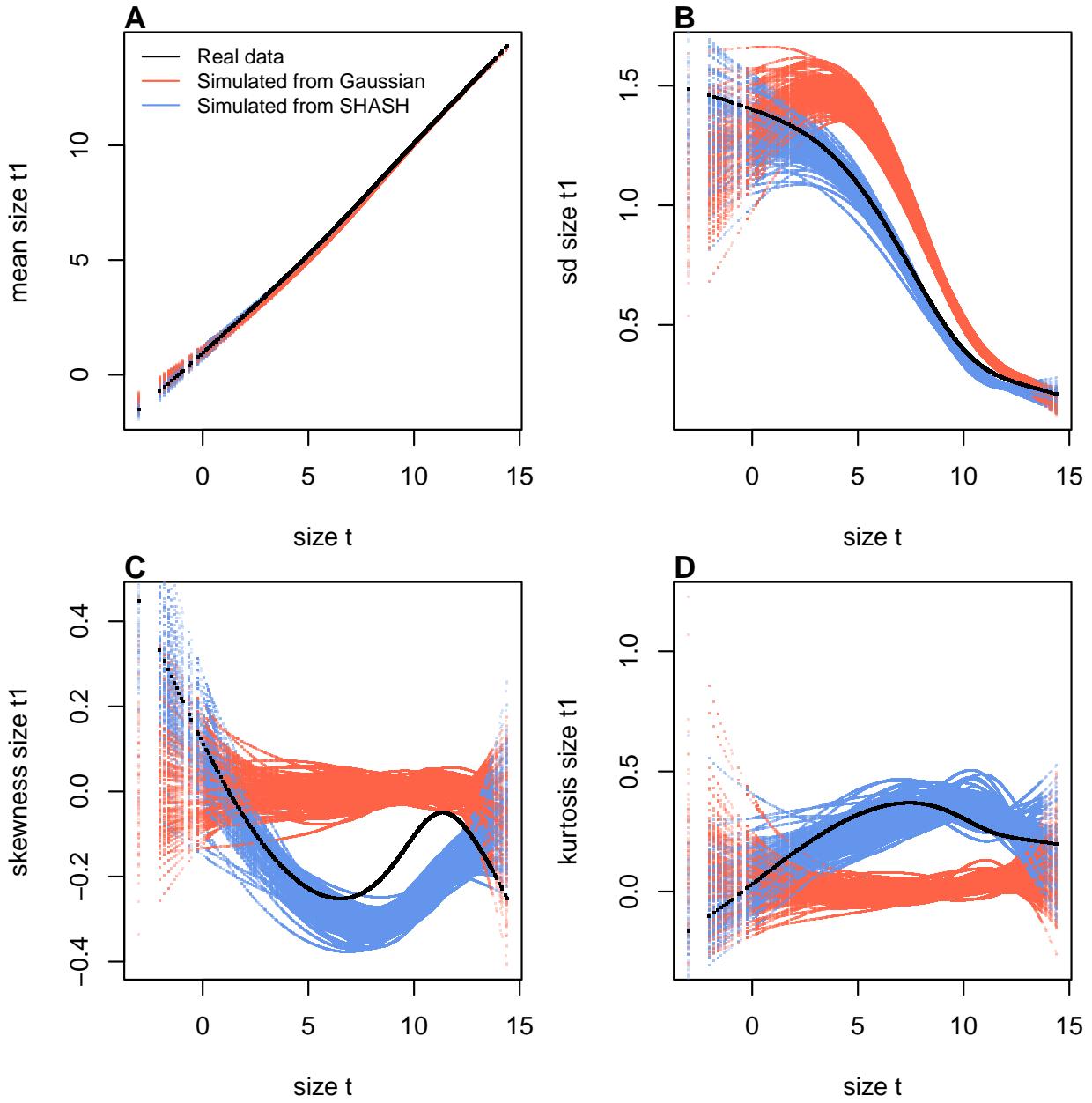


Figure 7: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

<sup>438</sup> of one's analysis, modeling non-Gaussian size transitions with a Gaussian growth model  
<sup>439</sup> could bias results in either direction, or make no difference at all.

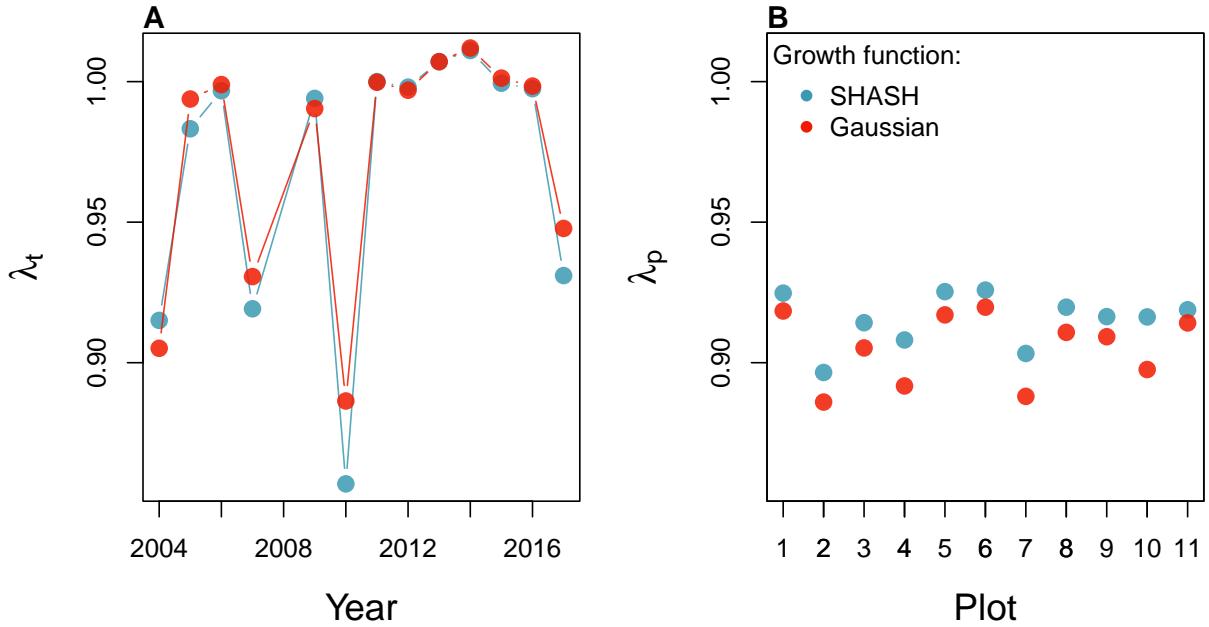


Figure 8: Temporal (A) and spatial (B) heterogeneity in fitness for the tree cholla cactus (*Cylindropuntia imbricata*) predicted by IPMs using Gaussian or SHASH growth models. Figure made by script `cactus_growth_modeling_qgam.R`.

### 440 3 Case study: creosotebush, *Larrea tridentata*

441 Our next case study comes from our studies of the woody shrub creosotebush (*Larrea tri-*  
 442 *dentata*) at the Sevilleta Long-Term Ecological Research (LTER) site in central New Mex-  
 443 ico, US. At this site as elsewhere in the Southwest US, creosotebush is encroaching into  
 444 desert grassland habitats. The data described here were collected along transects span-  
 445 ning grass-shrub ecotones to understand patterns of density dependence in creosotebush  
 446 demography. Specifically, we asked whether fitness is maximized approaching zero den-  
 447 sity at the leading edge of the expansion front (consistent with ‘pulled’ expansion), or  
 448 whether there is a demographic advantage for shrubs at higher density due to positive  
 449 feedbacks expected for ecosystem engineers (leading to ‘pushed’ expansion). Our pub-  
 450 lished study (Drees et al., 2023) used a spatial integral projection model (SIPM) to predict  
 451 the speed of shrub encroachment, assuming normally-distributed size transitions. Here  
 452 we step through our suggested workflow to ask whether a non-Gaussian model would  
 453 have been more faithful to the data, and how such an improvement would influence  
 454 predictions for the speed of encroachment. We use this case study to illustrate several  
 455 new elements and challenges, including modeling skewness and kurtosis as functions  
 456 of expected future size (instead of initial size) and using distributions that are not cur-

457 recently available as **mgcv** families. In fact, to diversify our use of software and illustrate  
458 alternatives, we do not use gam's for any element of this case study.

459 Growth data come from 522 shrubs censused longitudinally over four years (2013–  
460 2017). Census individuals occurred along 12 replicate transects (200 to 600 m in length)  
461 that spanned gradients of shrub density along shrub-grass ecotones. Size was measured  
462 as volume of an elliptical cone based on height and width measurements; the size vari-  
463 able of the IPM was the natural logarithm of volume ( $cm^3$ ). For each census individual,  
464 we recorded the size and density of all conspecifics within the five-meter transect “win-  
465 dow” in which it occurred, and took the sum of all sizes within the window as a measure  
466 of local density. The data are available in Ochocki et al. (2023).

467 As an initial Gaussian approach, we first fit a set of candidate generalized linear  
468 mixed models, including transect as a random effect, that represented competing hy-  
469 potheses for how size, density, and their interaction influence growth. Specifically, we fit  
470 five candidate Gaussian models that included fixed effects of initial size only (model 1),  
471 size and density (model 2), and size, density, and their interaction (model 3), allowing  
472 for shrubs of different sizes to have different growth responses to local density. Models  
473 4 and 5 mirrored models 2 and 3 but included second-order terms for density, allowing  
474 for the possibility of non-monotonic density dependence. As in (Drees et al., 2023) we  
475 pooled data across three transition years. Initial AIC rankings of these pilot models fa-  
476 vor model 4 slightly over model 5 ( $\Delta AIC = 0.8$ ) and significantly over all other models  
477 ( $\Delta AIC > 2$ ). However, these models were fit assuming constant variance, and inspection  
478 of the residuals of the best model indicate this is not a safe assumption.

479 Unlike our previous case studies, here we have multiple fixed effects that may influ-  
480 ence the variance of future size. In cases such as this, we recommend modeling variance  
481 as a function of expected future size rather than initial size, as we did with the corals  
482 and cacti. The expected (or “fitted”) values reflect the combined influence of all fixed  
483 and random effects, and therefore implicitly account for multiple sources of variation in  
484 the variance. While there are several convenient software packages for simultaneously  
485 modeling Gaussian mean and variance as functions of independent variables (**mgcv** for  
486 additive models as we saw above, **nlme** for linear models), **modeling variance as a func-**  
**487 *tion of the mean is trickier because they cannot easily be fit simultaneously***<sup>2</sup>. Here we  
488 us an iterative re-weighting approach – which is not elegant, but it works. For Gaus-  
489 sian models, weights  $w_i$  can be used to indicate that the observations  $y_i$  vary in their  
490 dispersion around the mean. In general, the iterative steps are:

---

<sup>2</sup>*After I wrote this I discovered that nlme can fit residual variance as a function of fitted(.)*.

1. Fit the expected value and normally-distributed residuals with constant variance:

$$y_i = \mu_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma)$$

2. Fit the standard deviation of the residuals as a function of the expected value.  
Weights are derived as the inverse of the fitted variance:

$$\epsilon_i \sim N(0, f(\mu_i))$$

$$w_i = 1/f(\mu_i)^2$$

3. Re-fit the observation model, weighting the residual variance according to step 2:

$$y_i = \mu_i + \epsilon_i$$

$$\epsilon_i \sim N(0, \sigma \times \sqrt{w_i})$$

491 We iterated steps 2 and 3 until the weights did not change. In step 2, we modeled  
 492 the standard deviation as a simple linear function of the expected value ( $\log(f(\mu_i)) =$   
 493  $\beta_0 + \beta_1 * \mu_i$ ) but other functions are possible, as is model selection among them. We  
 494 did this for all candidate models and, for fair AIC comparison, we re-fit all candidate  
 495 models with the same weights, estimated from the top model. The updated model  
 496 selection continued to favor model 4, but now with a stronger improvement over the  
 497 next-best model ( $\Delta AIC = 3.0$ ).

498 The resulting Gaussian growth model predicts strong initial size-dependence and  
 499 weak and slightly nonlinear (but monotonic) negative density dependence (Fig. 9A).  
 500 The model accounts for non-constant variance through the fitted weights, which indicate  
 501 greater dispersion for smaller values of expected size ( $\beta_1 = -0.21$ ; Fig. 9B). Quantiles of  
 502 the standardized residuals indicate weak negative skew (difference in tail size is 1–2%  
 503 of their total) and positive excess kurtosis, especially at smaller expected sizes (tails are  
 504 6–10% fatter than Gaussian) (Fig. 9C).<sup>3</sup> As a candidate for improvement, we turned to  
 505 the Johnson's  $S_U$  (JSU) distribution, a four-parameter, leptokurtic distribution capable  
 506 of skew in either direction. We used a parameterization of the JSU for which location

---

<sup>3</sup>Note that there is still a variance trend in the standardized residuals—rather unsatisfying! I have been through this backwards and forwards and my take is that this is a product of the sample size imbalance between small and large plants. The quantile regression is doing its best.

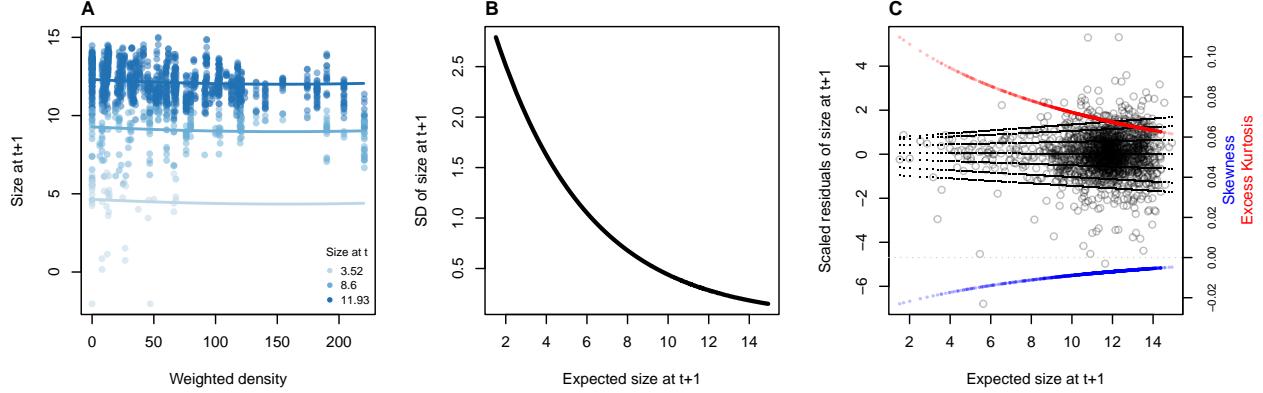


Figure 9: **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume ( $cm^3$ ). **B**, Standard deviation of size at time  $t + 1$  as a function of expected size at  $t + 1$  (the fitted values), estimated by iterative re-weighting. **C**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (blue) and excess kurtosis (red) derived from them. Black lines in **C** show the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles. Figure made by script `creosote_growth_modeling_qgam.R`.

parameter  $\mu$  is the mean and scale parameter  $\sigma$  is the standard deviation (Rigby et al., 2019).

Like many of the non-Gaussian candidates that we suggest (Fig. 1), the JSU distribution is not presently available as a family option for linear mixed models in any software package, to our knowledge. However, this need not be a barrier to using it for growth modeling. We fit a custom maximum likelihood model that borrows the mean and standard deviation of best Gaussian model and limits estimation of free parameters to those that control the JSU's skewness and kurtosis – effectively modeling the standardized residuals rather than sizes. Here is what such a hybrid likelihood model looks like in practice:

```

517  ## log_volume_t1 are the size obervations
518  ## GAU_fitted are the expected values of the best Gaussian model
519  ## pars is a vector of free parameters to be estimated
520  JSULogLik=function(pars){
521    dJSU(x=log_volume_t1,
522          mu=GAU_fitted,
523          sigma=exp(GAU_sd_coef[1]+GAU_sd_coef[2]*GAU_fitted),
524          nu = pars[1]+pars[2]*GAU_fitted,
525          tau = exp(pars[3]+pars[4]*GAU_fitted), log=TRUE)

```

526 }

527 The mean of the JSU is set to that of the best Gaussian model (GAU\_fitted) and the  
528 standard deviation is a function of the mean according to the coefficients (GAU\_sd\_coef)  
529 estimated through iterative re-weighting. Based on diagnostics of the standardized resid-  
530 uals (Fig. 9), JSU parameters that control skewness and kurtosis are defined as linear  
531 functions of the mean, and it is these coefficients that are estimated by maximum like-  
532 lihood. Here we are relying on the robustness of Gaussian linear models to deviations  
533 from normality . If one is skeptical of this approach, it is possible, as an alternative,  
534 to simultaneously re-fit all parameters of the JSU in a maximum likelihood framework.  
535 However, incorporating random effects into a custom likelihood model is non-trivial (we  
536 provide guidance on one way to do this, using the “shrinkage” approach, in Appendix  
537 XX). Therefore a key advantage of the hybrid approach is convenient retention of the  
538 fitted random effects and associated variance components, which get shuttled from the  
539 Gaussian model into the non-Gaussian model without any fuss (it was critical that we  
540 used a parameterization of the JSU for which `mu` is the mean and `sigma` is the standard  
541 deviation). And, if this approach does not “work” (i.e., deviations from normality bi-  
542 ased the fitted values of the Gaussian model) one would quickly find out through the  
543 simulation step of the workflow. In this case, the hybrid JSU model performed well,  
544 generating simulated data that aligned with the real data better than the best Gaussian  
545 model, particularly in **standard deviation**<sup>4</sup> and kurtosis (Fig. 10). Note that in Fig. 10  
546 we are plotting moments of the future size distribution with respect to initial size; this  
547 distribution is also conditional on density but initial size is by far the stronger predictor  
548 of future size, so we chose this visualization.

549 The improvement of the JSU over the Gaussian growth model, while visually satis-  
550 fying, had virtually no influence on SIPM results. Models using Gaussian or JSU growth  
551 kernels had nearly identical, monotonic decreases in  $\lambda$  with increasing local density, and  
552 nearly identical wave velocities (Fig. 11). This species has very low mortality risk once  
553 established (mean remaining life expectancy of a median-sized shrub is 24,408 years)  
554 and its population growth and wave expansion are limited by very low seedling recruit-  
555 ment ((Drees et al., 2023)). Weak size-dependence in survival likely explains why the  
556 improvement in growth modeling had little influence on SIPM predictions.

---

<sup>4</sup> *I am a little mystified as to why the JSU is so much better. It is literally the same SD in both distributions.*

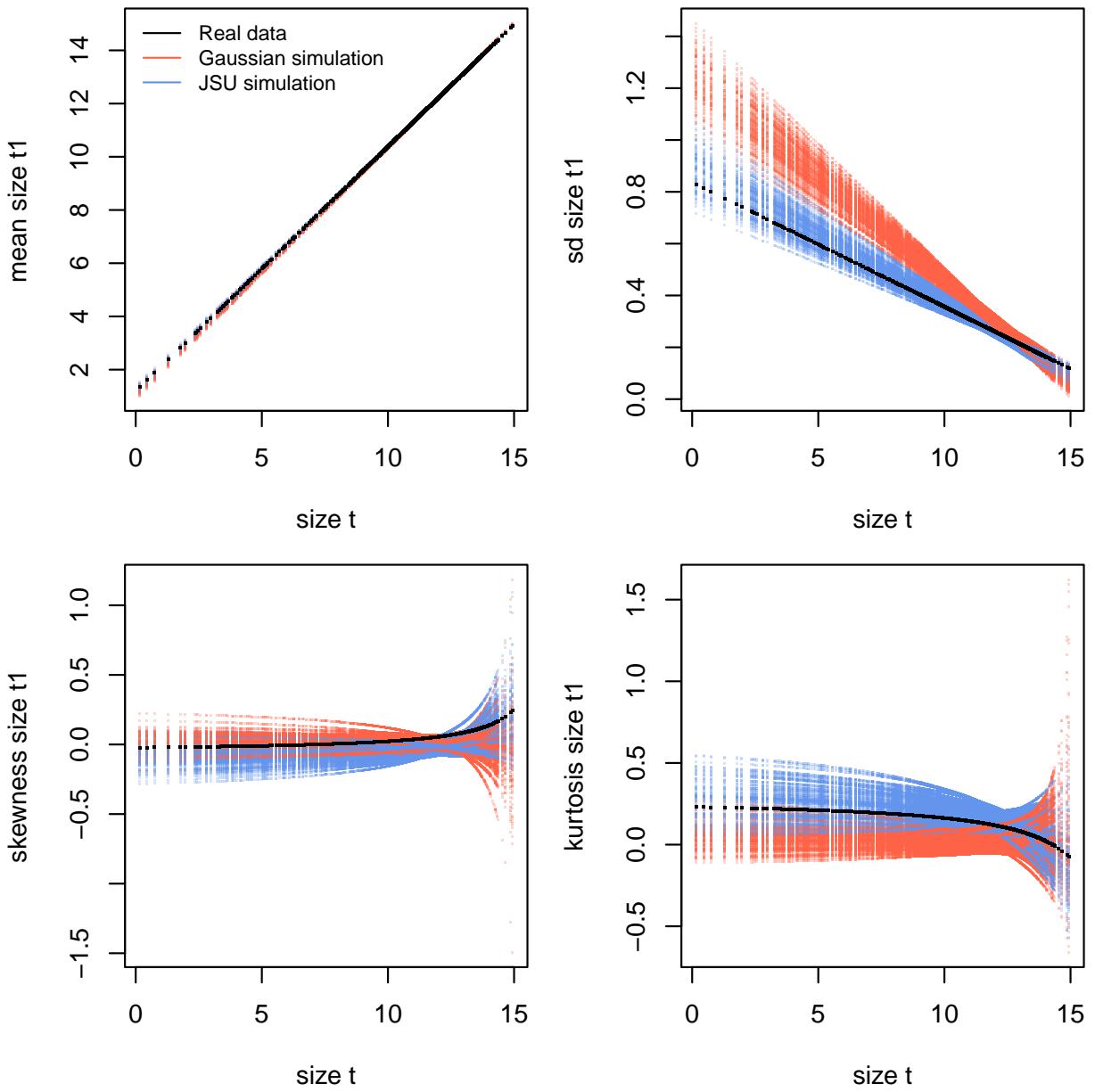


Figure 10: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `creosote_growth_modeling_qgam.R`.

## 557 4 Case study: lady orchid, *Orchis purpurea*

558 Our final case study examines selection on life history strategies in the lady orchid *Or-*  
 559 *chis purpurea*. In a prior study, Miller et al. 2012 contrasted the growth trajectories from  
 560 year  $t$  to  $t + 1$  for plants that did or did not flower in year  $t$ , as a way to quantify costs

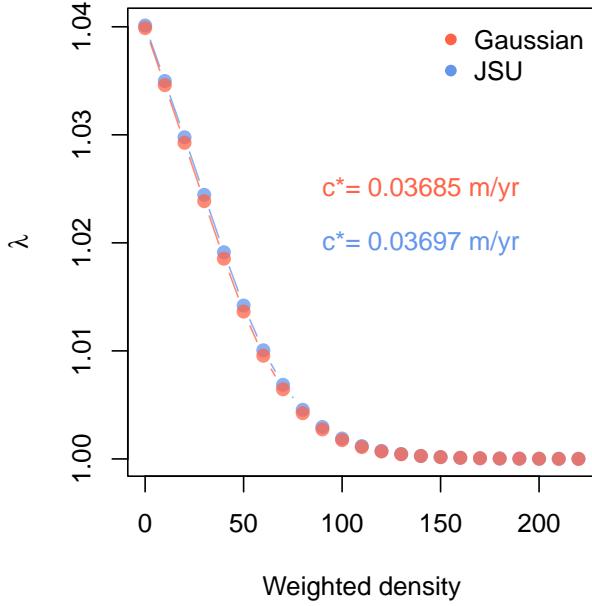


Figure 11: Density dependence in fitness ( $\lambda$ ) and asymptotic velocity of the creosote encroachment wave ( $c^*$ ) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ( $\log(cm^3)$ ) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

of reproduction. The different growth kernels were then used in an IPM to quantify evolutionarily stable life history strategies: the optimal flowering size that balances benefits of flowering at larger sizes against the risk of dying before reaching those sizes. The original study assumed a Gaussian distribution of size transitions and allowed for non-constant variance with respect to initial size. Here we re-visit that analysis applying our growth modeling workflow to derive improved growth kernels for flowering and non-flowering orchids.

The data, originated by Dr. Hans Jacquemyn and used here with permission, come from 368 plants in a Belgian population that was censused annually from 2003 through 2011 (for this reanalysis we are using data only from the “light” habitat). Size was measured as leaf area ( $cm^3$ ) summed over all leaves, and we analyzed the natural logarithm of total leaf area as the size variable of the IPM.

As a pilot Gaussian approach, we fit six candidate models in which the mean was a function of initial size only, additive effects of initial size and flowering status, and interaction between size and flowering, and the standard deviation was a function of size only (models 1-3) or size and flowering status (models 4-6). All models included a random intercept for year. As another variation on software and an alternative to two-

578 step fitting or iterative re-weighting, here we use `nmle::lme()`, which can simultaneously  
579 fit linear predictors for mean and variance. For example, model 1 was:

```
580 orchid_GAU[[1]]<-lme(log_area_t1~ log_area_t,  
581 weights=varExp(form=~ log_area_t),  
582 random=~ 1|begin.year,data=orchid_grow,method="ML")
```

583 Model 3 (size  $\times$  flowering) was strongly favored, consistent with prior results that non-  
584 flowering plants have a growth advantage over flowering plants. Growth variance de-  
585 clined with initial size for both reproductive classes (Fig. 12A-B) and skewness and kur-  
586 tosis of the standardized residuals indicate strong deviations from normality (Fig. 12C-  
587 D). For most sizes, left skew and excess kurtosis were more severe for non-reproductive  
588 plants, with tail imbalance ca. 10% of their total and tail weights 10–20% fatter than  
589 Gaussian.

590 As improvements, we explored the skewed *t* and Johnson's SU distributions, both  
591 leptokurtic distributions with flexible skewness. We were happier with the skewed *t*,  
592 which we fit in a similar way as we fit the JSU to the creosote data, setting the mean  
593 and standard deviation to the Gaussian fits and estimating free parameters controlling  
594 skewness and kurtosis:

```
595 ## log_area_t1 and log_area_t are the size obervations  
596 ## flowering indicates reproductive status at time t (0 or 1)  
597 ## GAU_fitted and GAU_sd are mean and standard deviation from lme  
598 ## pars is a vector of free parameters to be estimated  
599 SSTLogLik=function(pars){  
600     dSST(x=log_area_t1,  
601             mu=GAU_fitted,  
602             sigma=GAU_sd,  
603             nu = exp(pars[1] + pars[2]*log_area_t + pars[3]*as.logical(flowering) + pars[4])  
604             tau = exp(pars[5] + pars[6]*log_area_t + pars[7]*as.logical(flowering) + pars[8])  
605             log=TRUE)  
606 }
```

607 `gamlss.dist:dSST` is a parameterization of the skewed *t* in which `mu` and `sigma` are the  
608 mean and standard deviation, respectively. Based on diagnostics of the standardized  
609 residuals (Fig. 12) we allowed `nu` and `tau` to vary by size and differ between flowering  
610 and non-flowering plants (note that the `tau` parameter uses a  $\log(x - 2)$  link function).  
611 Size transition data simulated from this model corresponded favorably to the real data,

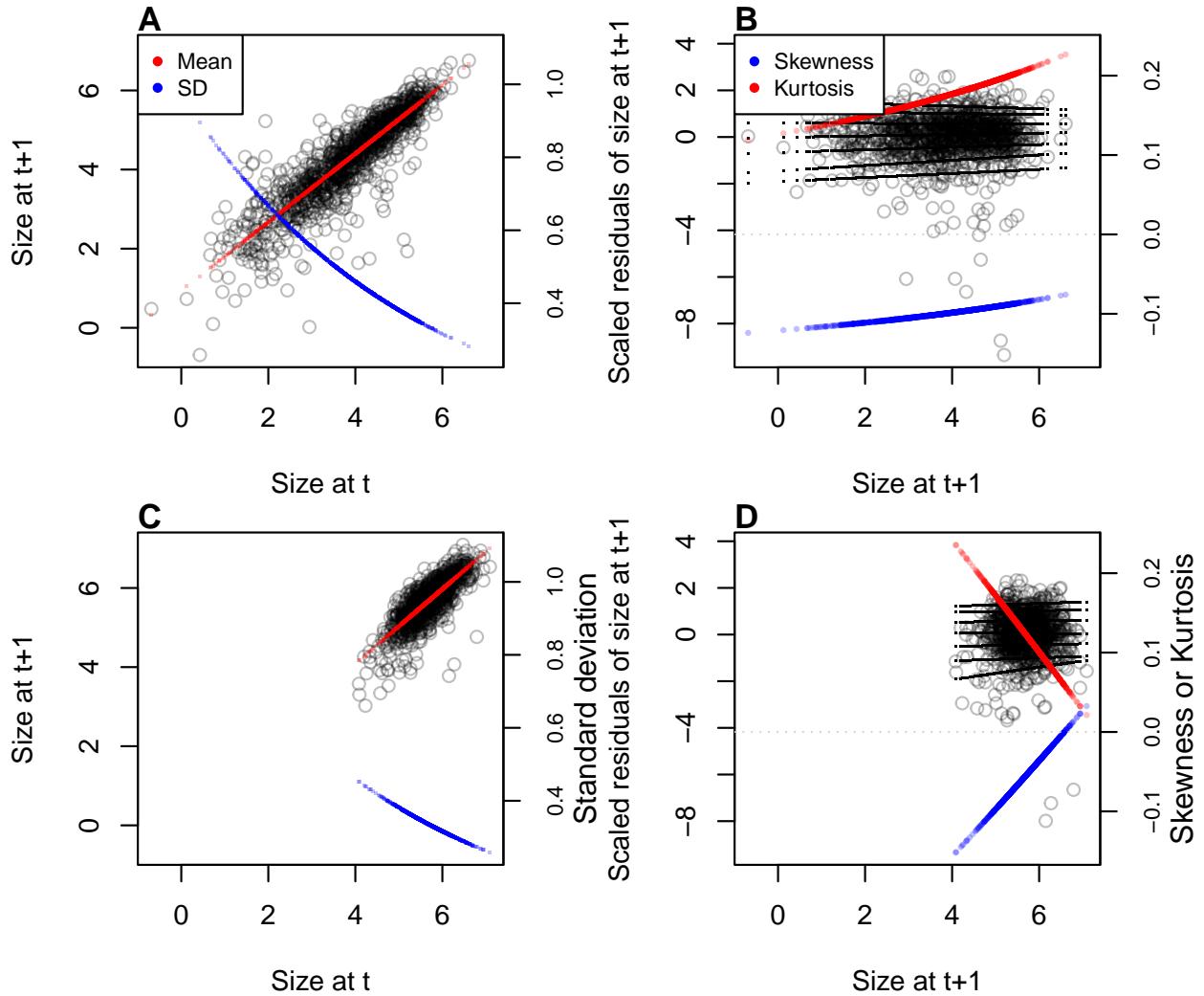


Figure 12:

much better than the pilot Gaussian model, including improvements in the **standard deviation**<sup>5</sup>, skewness, and kurtosis of future size (Fig. 13).

Finally, we used the improved growth model to revisit key results of the original study. Miller et al. (2012) used the orchid IPM to estimate the evolutionarily stable strategy (ESS) as the mean size at flowering that maximizes lifetime reproductive success ( $R_0$ ), given the constraint that flowering when small reduces growth and thus elevates mortality risk. Repeating that analysis here, we found that improved growth modeling has virtually no influence on predictions for optimal life history strategies (Fig. 14). ESS flowering sizes were nearly identical between IPMs with Gaussian vs skewed  $t$  growth models, and both aligned well with the observed mean flowering size (dashed vertical

<sup>5</sup>Again, the improvement here is surprising to me and I am unsure what to say about it.

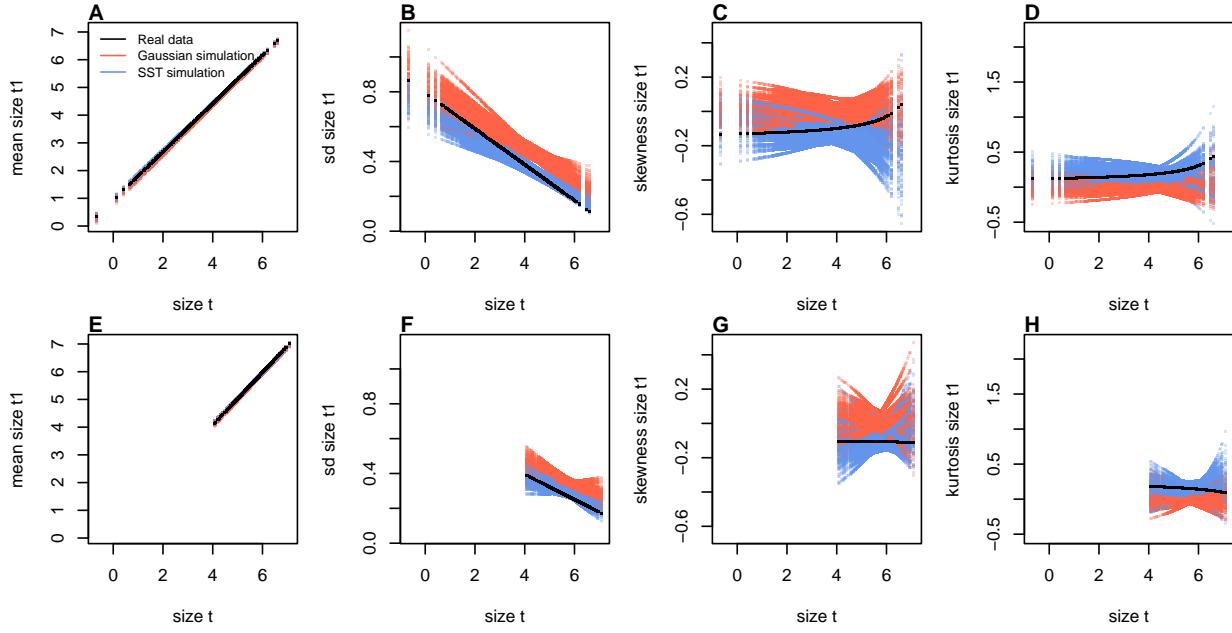


Figure 13: Comparisons between real orchid data and data simulated from Gaussian and skewed  $t$  growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (A-D) shows plants that were vegetative at the start of the transition year and bottom row (E-H) shows plants that were flowering at the start of the transition year. Figure made by script `orchid_growth_modeling_rq.R`.

line in Fig. 14A). Extending beyond the original study, we also explored expected remaining lifespan for different ages and sizes (R package **Rage** (Jones et al., 2022)). Gaussian and skewed  $t$  growth models predicted nearly identical mean remaining lifespans across the stage and size distribution (Fig. 14B). However, the skewed  $t$  model predicted consistently greater variance in remaining lifespan, nearly 10% greater at some sizes.<sup>6</sup> Thus, as we have seen in other case studies, the practical consequences of improved growth modeling depend on what one aims to learn from the IPM.

## 5 Discussion

Much of the appeal of integral projection models has stemmed from their embrace of continuous size structure through reliance on regression-based approaches, and the potentially complex fixed- and random-effect structures that these approaches allow. Using familiar statistical tools and with relatively few parameters to estimate, IPM users can

<sup>6</sup>Do not believe this result! I have left it here as a placeholder because I would like to do this correctly. But I think there are problems with `Rage's life_expect_var()` function. The predicted variance declines linearly with matrix dimension.

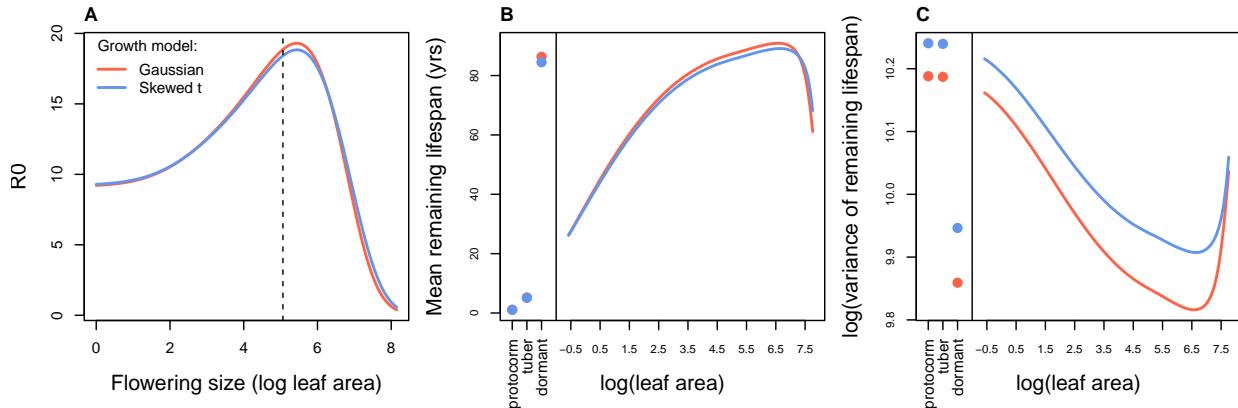


Figure 14: Orchid life history results from IPMs using Gaussian or skewed  $t$  growth models. **A**, Lifetime reproductive success ( $R_0$ ) as a function of mean size of flowering. Dashed vertical line shows the observed mean flowering size. **B-C**, Mean and variance of remaining lifespan as a function of size or stage. The orchid IPM includes three discrete below-ground stages (protocorm, tuber, and dormant plant) in addition to continuous size of above-ground plants.

incorporate important sources of variation in demography and interrogate their influence on ecological and evolutionary dynamics. With this opportunity comes the burden of getting it right: IPMs are good models of the populations they are intended represent only insofar as the statistical models provide good fits to the underlying data. The growth sub-model is the trickiest part of “getting it right” because it defines a distribution of future size conditional on current size. Distributions have many properties – “moments” – and a good growth model should recapitulate the properties of real size transitions. The default assumption of normally distributed size transitions, employed overwhelmingly across 20+ years of IPM studies, is an arbitrary historical precedent. In four case studies (chosen simply because we had the data at our fingertips) and, we suspect, more broadly, skewness and excess kurtosis were common features of size transition data: shrinking was more common than growing, and large changes in size were more common than a Gaussian model would predict. Our most important message is that the standard assumption of normally-distributed size transitions should be abandoned and a more inquisitive process of growth modeling should take its place.

We have attempted to lay out a general workflow for what that process should look like, guided by visual diagnostics of standardized residuals. One implication of relying on visual diagnostics is that goodness of fit is in the eye of the beholder. This approach can empower IPM users to make informed choices, but it is not very prescriptive: we have not suggested any hard rules for when one or another distribution should be used,

654 only that a good growth model should generate data that look like the real thing. Al-  
655 ternatively, model selection could be used to identify best-fitting growth distributions  
656 and best-fitting functions for higher moments. However, model selection among growth  
657 distributions with 3-5 parameters, each of which may be functions of state variables or  
658 fitted values, can quickly explode in complexity, and we are not convinced it is worth  
659 the trouble. It is possible to find a good growth model without worrying about which  
660 one is “best”.

661 In all of our case studies, non-Gaussian growth models always yielded more sat-  
662 isfying fits to size transition data than the Gaussian models published in those papers.  
663 However, much to our relief, none of these re-analyses yielded a “gotcha” result that  
664 overturned results of the original study. In this small sampling of case studies, im-  
665 proved growth modeling had only modest effects on IPM results. We caution against  
666 taking too much comfort in this outcome; we can imagine other scenarios in which the  
667 choice of the growth distribution could be more consequential. It is worth noting that  
668 three of our case studies focused on perennial plants and the fourth focused on corals,  
669 which are demographically similar to perennial plants (heavy losses during recruitment  
670 but high survival once established). Life cycles such as these may be relatively robust to  
671 subtle features of the growth kernel. More systematic comparative analyses across may  
672 provide insight into which types of species and life histories are more likely to exhibit  
673 strong skewness and kurtosis of size transitions, and the conditions under which demo-  
674 graphic analysis is more or less sensitive to these features of size transition. It is also  
675 worth noting, as we saw in several case studies, that different outputs from the same  
676 model can be more or less sensitive to the choice of growth distribution.

677 Some issues to be discussed.

- 678 • Many software options: lme4/maxLik, mgcv, rstan
- 679 • Comparison of our approach with beta regression method.
- 680 • We have emphasize growth but same principles apply to other continuous state  
681 transitions, eg disease IPMs.

## 682 Acknowledgements

683 This research was supported by US NSF grants DEB-1933497 to SPE and DEB-1754468,  
684 2208857, and 2225027 to TEXM.

685 **6 Authorship statement**

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

688 **7 Data accessibility statement**

689 No original data appear in this paper. Should the paper be accepted, all computer scripts  
690 supporting the results will be archived in a Zenodo package, with the DOI included at  
691 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).  
692

## 693 Literature Cited

- 694 Anscombe, F. J. and Glynn, W. J. (1983). Distribution of the kurtosis statistic  $b_2$  for  
695 normal samples. *Biometrika*, 70(1):227–234.
- 696 Bates, D., Sarkar, D., Bates, M. D., and Matrix, L. (2007). The lme4 package. *R package  
697 version*, 2(1):74.
- 698 Bruno, J. F., Ellner, S. P., Vu, I., Kim, K., and Harvell, C. D. (2011). Impacts of aspergillosis  
699 on sea fan coral demography: modeling a moving target. *Ecological Monographs*,  
700 81(1):123–139.
- 701 Compagnoni, A., Bibian, A. J., Ochocki, B. M., Rogers, H. S., Schultz, E. L., Sneck, M. E.,  
702 Elderd, B. D., Iler, A. M., Inouye, D. W., Jacquemyn, H., et al. (2016). The effect of  
703 demographic correlations on the stochastic population dynamics of perennial plants.  
704 *Ecological Monographs*, 86(4):480–494.
- 705 Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide  
706 to bayesian model checking for ecologists. *Ecological Monographs*, 88(4):526–542.
- 707 Cooch, E. G. and White, G. C. (2020, accessed 5/17/2020). *Program MARK - a 'gentle  
708 introduction'*. Available at phidot.org.
- 709 Coulson, T. (2012). Integral projections models, their construction and use in posing  
710 hypotheses in ecology. *Oikos*, 121(9):1337–1350.
- 711 Crone, E. E. (2016). Contrasting effects of spatial heterogeneity and environmental  
712 stochasticity on population dynamics of a perennial wildflower. *Journal of Ecology*,  
713 104(2):281–291.
- 714 Czachura, K. and Miller, T. E. (2020). Demographic back-casting reveals that subtle  
715 dimensions of climate change have strong effects on population viability. *Journal of  
716 Ecology*.
- 717 D'Agostino, R. B. (1970). Transformation to normality of the null distribution of  $g_1$ .  
718 *Biometrika*, pages 679–681.
- 719 Davis, C. (2015). *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.
- 720 Drees, T., Ochocki, B. M., Collins, S. L., and Miller, T. E. (2023). Demography and  
721 dispersal at a grass-shrub ecotone: a spatial integral projection model for woody plant  
722 encroachment. *Ecological Monographs*, page e1574.

- 723 Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying  
724 a new structured population model. *Ecology*, 81(3):694–708.
- 725 Elderd, B. D. and Miller, T. E. (2016). Quantifying demographic uncertainty: Bayesian  
726 methods for integral projection models. *Ecological Monographs*, 86(1):125–144.
- 727 Ellis, M. M. and Crone, E. E. (2013). The role of transient dynamics in stochastic popula-  
728 tion growth for nine perennial plants. *Ecology*, 94(8):1681–1686.
- 729 Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E., and Rees, M. (2022).  
730 A critical comparison of integral projection and matrix projection models for demo-  
731 graphic analysis: Comment. *Ecology*, 103(10):e3605.
- 732 Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modeling of Structured Popula-  
733 tions: A Practical Guide to the Integral Projection Model*. Springer, New York.
- 734 Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in  
735 animal populations. *Ecology*, 79:2531 – 2538.
- 736 Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New  
737 York, 2 edition.
- 738 Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed  
739 models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.
- 740 Héault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C. T., Wagner,  
741 F., and Baraloto, C. (2011). Functional traits shape ontogenetic growth trajectories of  
742 rain forest tree species. *Journal of ecology*, 99(6):1431–1440.
- 743 Jones, M. and Pewsey, A. (2009). *Biometrika*, 96:761 – 780.
- 744 Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis.  
745 *The American Statistician*, 65(2):89 – 95.
- 746 Jones, O. R., Barks, P., Stott, I., James, T. D., Levin, S., Petry, W. K., Capdevila, P., Che-  
747 Castaldo, J., Jackson, J., Römer, G., et al. (2022). Rcompadre and rage—two r pack-  
748 ages to facilitate the use of the compadre and comadre databases and calculation of  
749 life-history traits from matrix population models. *Methods in Ecology and Evolution*,  
750 13(4):770–781.
- 751 Komsta, L. and Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and  
752 related tests. *R package version*, 14(1).

- 753 Link, W. A. and Nichols, J. D. (1994). On the importance of sampling variance to inves-  
754 tigations of temporal variation in animal population size. *Oikos*, 69(3):539 – 544.
- 755 Louthan, A. M., Keighron, M., Kiekebusch, E., Cayton, H., Terando, A., and Morris, W. F.  
756 (2022). Climate change weakens the impact of disturbance interval on the growth rate  
757 of natural populations of venus flytrap. *Ecological Monographs*, 92(4):e1528.
- 758 McGillivray, H. (1986). Skewness and asymmetry: measures and orderings. *Annals of  
759 Statistics*, 14:994–1011.
- 760 Metcalf, C. J. E., Ellner, S. P., Childs, D. Z., Salguero-Gómez, R., Merow, C., McMahon,  
761 S. M., Jongejans, E., and Rees, M. (2015). Statistical modelling of annual variation for  
762 inference on stochastic population dynamics using Integral Projection Models. *Methods  
763 in Ecology and Evolution*, 6:1007–1017.
- 764 Miller, T. E. (2020). Long-term study of tree cholla demography in the los pinos  
765 mountains, sevilleta national wildlife refuge. [https://doi.org/10.6073/pasta/  
766 dd06df3f950afe4a4642306182237d13](https://doi.org/10.6073/pasta/dd06df3f950afe4a4642306182237d13).
- 767 Miller, T. E., Louda, S. M., Rose, K. A., and Eckberg, J. O. (2009). Impacts of insect  
768 herbivory on cactus population dynamics: experimental demography across an envi-  
769 ronmental gradient. *Ecological Monographs*, 79(1):155–172.
- 770 Miller, T. E., Williams, J. L., Jongejans, E., Brys, R., and Jacquemyn, H. (2012). Evolution-  
771 ary demography of iteroparous plants: incorporating non-lethal costs of reproduction  
772 into integral projection models. *Proceedings of the Royal Society B: Biological Sciences*,  
773 279(1739):2831–2840.
- 774 Ochocki, B. M., Drees, T., and Miller, T. E. (2023). Density-dependent demography of  
775 creosote bush (*larrea tridentata*) along grass-shrub ecotones. <https://doi.org/10.6073/pasta/ca53c16f16dcf9fb11f3ee99ea5445ac>.
- 777 Ohm, J. R. and Miller, T. E. (2014). Balancing anti-herbivore benefits and anti-pollinator  
778 costs of defensive mutualists. *Ecology*, 95(10):2924–2935.
- 779 Ozgul, A., Childs, D. Z., Oli, M. K., Armitage, K. B., Blumstein, D. T., Olson, L. E., Tul-  
780 japurkar, S., and Coulson, T. (2010). Coupled dynamics of body mass and population  
781 growth in response to environmental change. *Nature*, 466(7305):482–485.

- 782 Peterson, M. L., Morris, W., Linares, C., and Doak, D. (2019). Improving structured  
783 population models with more realistic representations of non-normal growth. *Methods*  
784 in Ecology and Evolution, 10(9):1431–1444.
- 785 Plard, F., Schindler, S., Arlettaz, R., and Schaub, M. (2018). Sex-specific heterogeneity  
786 in fixed morphological traits influences individual fitness in a monogamous bird  
787 population. *The American Naturalist*, 191(1):106–119.
- 788 Rees, M., Childs, D. Z., and Ellner, S. P. (2014). Building integral projection models: a  
789 user's guide. *Journal of Animal Ecology*, 83(3):528–545.
- 790 Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., and De Bastiani, F. (2019). *Distributions*  
791 for modeling location, scale, and shape: Using GAMLSS in R. CRC press.
- 792 Salguero-Gómez, R. and Casper, B. B. (2010). Keeping plant shrinkage in the demo-  
793 graphic loop. *Journal of Ecology*, 98(2):312–323.
- 794 Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky,  
795 C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020).  
796 Robustness of linear mixed-effects models to violations of distributional assumptions.  
797 *Methods in ecology and evolution*, 11(9):1141–1152.
- 798 Schultz, E. L., Eckberg, J. O., Berg, S. S., Louda, S. M., and Miller, T. E. (2017). Native  
799 insect herbivory overwhelms context dependence to limit complex invasion dynamics  
800 of exotic weeds. *Ecology letters*, 20(11):1374–1384.
- 801 Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location  
802 scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46.
- 803 Stubberud, M. W., Vindenes, Y., Vøllestad, L. A., Winfield, I. J., Stenseth, N. C., and Lan-  
804 gangen, Ø. (2019). Effects of size-and sex-selective harvesting: An integral projection  
805 model approach. *Ecology and Evolution*, 9(22):12556–12570.
- 806 Wan, X., Wang, W., Liu, J., and Tong, T. (2014). Estimating the sample mean and stan-  
807 dard deviation from the sample size, median, range and/or interquartile range. *BMC*  
808 *medical research methodology*, 14:1–13.
- 809 Williams, J. L., Miller, T. E., and Ellner, S. P. (2012). Avoiding unintentional eviction from  
810 integral projection models. *Ecology*, 93(9):2008–2014.

811 Wood, S. (2017). *Generalized Additive Models: An Introduction with R.* Chapman and  
812 Hall/CRC, 2 edition.

813 Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational  
814 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

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

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

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

853 
$$X_{\lambda, \tau} = \sinh(e^{-\tau} \operatorname{asinh}(Z) + \lambda). \quad (\text{S.3})$$

854 Thus, the two parameterizations are related by

855 
$$\delta = e^\tau, \epsilon = \delta\lambda = e^\tau\lambda. \quad (\text{S.4})$$

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

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

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

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

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

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

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

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

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

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

904 If there are  $T$  years, the matrix `U` specified in this way has  $2T$  columns corresponding to  
905  $n$  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:

```

909 LogLik=function(pars,new.size,U){
910   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
911   mu = U%*%pars1;
912   sigma = exp(pars2[1]+pars2[2]*mu);
913   dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)
914 }
```

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

```

917 start=c(runif(ncol(U)), rep(0,3))
918 out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,
919   method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);
920 coefs = out$estimate; # parameters
921 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

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

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

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

$$946 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

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

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

$$954 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

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

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

$$960 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

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

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

$$968 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.10})$$

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

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

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

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

```
989 # Variance-covariance matrices for intercepts and slopes
990 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
991 # Extract year-specific intercepts, center them to zero
992 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
993
994 # Estimate sigma^2
995 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
996           (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
997
998 # Shrink deviations from the mean
```

```

999 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
1000
1001 # Do it all again for the slopes
1002 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
1003 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
1004 (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
1005 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

1006 The figure below shows the results for one artificial PSSP “data” set, having  $T = 22$   
 1007 years and growth measurements on about 175 individuals/year on average. The true  
 1008 random year effects (the ones used to generate the data) are recovered with good accu-  
 1009 racy and no bias. In particular there is no sign of extreme values being pulled in too  
 1010 far towards the mean, which would cause an S-shaped graph of estimated versus true  
 1011 values.

