

# 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 without sacrificing ecological  
<sup>25</sup> complexity; our worked examples demonstrate how, including open-access data and  
<sup>26</sup> computing scripts. Doing so, as guided by careful interrogation of the data, will re-  
<sup>27</sup> sult in a model that better represents the 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 via-  
32 bility that are widely used in basic and applied settings. In contrast to MPMs for pop-  
33 ulations with discrete structure (life stage, age class, etc.), IPMs (Easterling et al., 2000)  
34 readily accommodate populations structured by continuous state variables, most com-  
35 monly size. A related innovation of the IPM framework is its emphasis on regression-  
36 based modeling for parameter estimation, which often carries important advantages for  
37 making the most of hard-won data (Ellner et al., 2022).

38 A standard workflow allows ecologists to assemble an IPM from data using famili-  
39 iar 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 only need to provide a single mean value 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.). But for  
54 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-

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

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

80 The persistence of Gaussian growth modeling is understandable. There is a long  
81 tradition of statistical modeling built on the assumption of normally distributed resid-  
82 uals with constant variance. Popular packages such as `lme4` (Bates et al., 2007), `mgcv`  
83 (Wood, 2017), and `MCMCglmm` (Hadfield et al., 2010) make it easy to fit growth models  
84 with potentially complex fixed- and random-effect structures, but the possible distribu-  
85 tions of continuous responses are limited, and default to Gaussian. Abandoning these  
86 convenient tools for the sake of more flexible growth modeling means, it may seem,  
87 sacrificing the flexibility to rigorously model diverse and potentially complex sources of  
88 variation in growth, some of which may be the motivation driving the study in the first  
89 place.

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

94 Our goal here is to present and illustrate a general and practical "recipe" that moves  
95 growth modeling past the standards set over 20 years ago. Like any recipe, users may  
96 need to make substitutions or add ingredients to suit their situation. Our approach  
97 emphasizes graphical diagnostics for developing and evaluating growth models, rather  
98 than a process centered on statistical model selection. Through a set of empirical case  
99 studies we demonstrate how a simple workflow, using tools that were nonexistent or not

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

## 108 A general workflow for better growth modeling

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

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

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

129 Allowing non-constant variance means that it is not necessary to transform the  
130 data in a way that stabilizes the growth variance. Transformation remains an option  
131 when it does not create new problems (see Discussion), and it may have advantages

132 besides variance stabilization. In particular log-transformation is often appropriate  
133 for size data (Ellner et al., 2016), and it helps to avoid eviction at small sizes.

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

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

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

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

165 In models with multiple covariates and/or random effects, each potentially affecting  
166 several distribution parameters (location, scale, skew, kurtosis) in different ways,

“refit the model” could entail a massive model selection process to identify the “right” or “best” non-Gaussian model. And with so many options, model uncertainty 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. Our recommendation for this step is based on the fact that parameter estimation using Gaussian regression models 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 a full-blown model selection exercise. Instead, alternative models should be chosen to remedy observable discrepancies between real and simulated size transition data, and at most slightly modified based on final diagnostics and statistical tests.

## How should skewness and kurtosis be measured?

Improvement of a Gaussian model will 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,

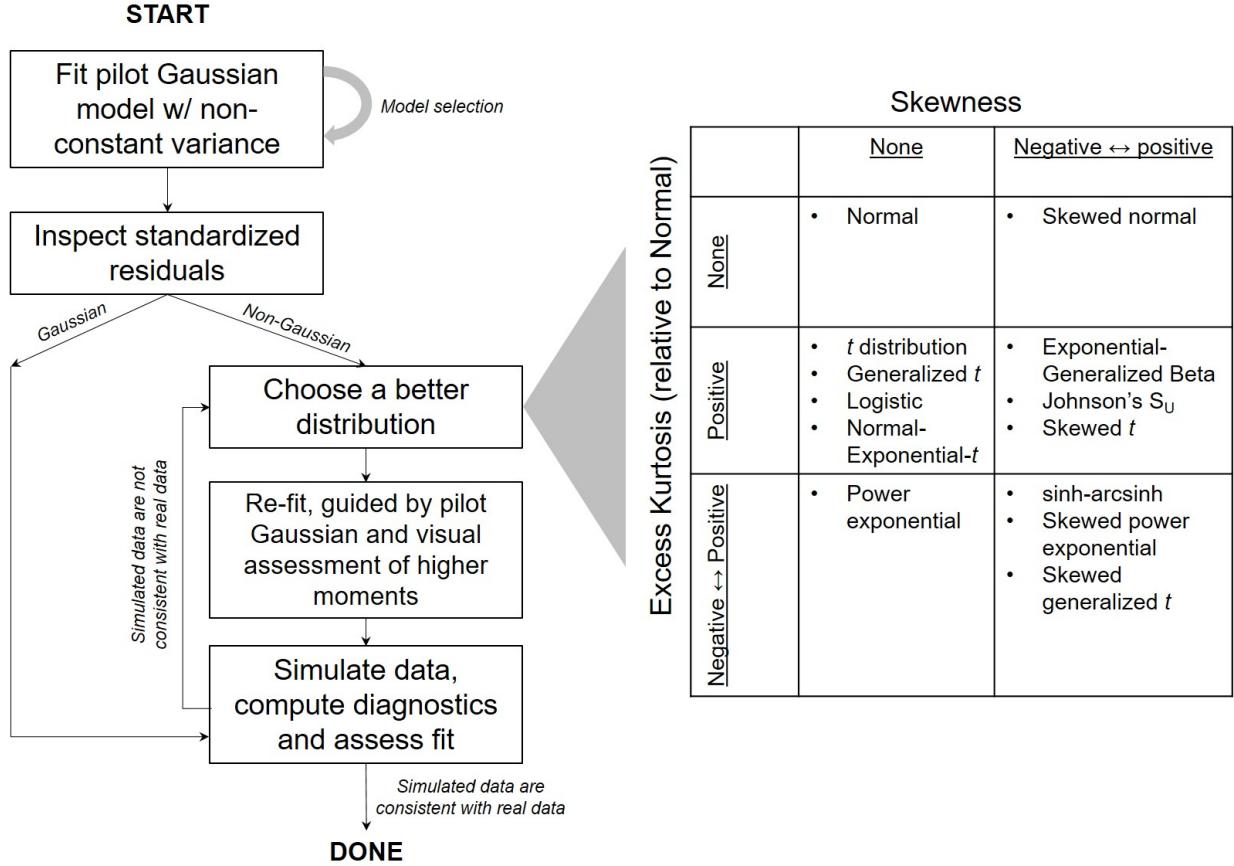


Figure 1: General workflow of recommendations for IPM growth modeling (left) and guide to common non-Gaussian distributions of size  $x$  for  $x \in \mathbb{R}$  that can accommodate different combinations of skewness and kurtosis (right). All of these distributions (often including multiple versions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

200 respectively, of the distribution:

$$201 \quad \text{Skewness} = \frac{m_3}{\sigma^3}, \quad \text{Excess kurtosis} = \frac{m_4}{\sigma^4} - 3 \quad (1)$$

202 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  
 203 variance (second central moment). A Gaussian distribution has zero skewness and zero  
 204 excess kurtosis.

205 The standard measures are easy to calculate but their use for choosing and eval-  
 206 uating growth models is hindered by their poor sampling properties. Because empirical  
 207 estimates involve high powers of data values, a few outliers can produce very inaccurate  
 208 estimates. Figure 2 shows a simulated example, where the underlying “data” are a sam-

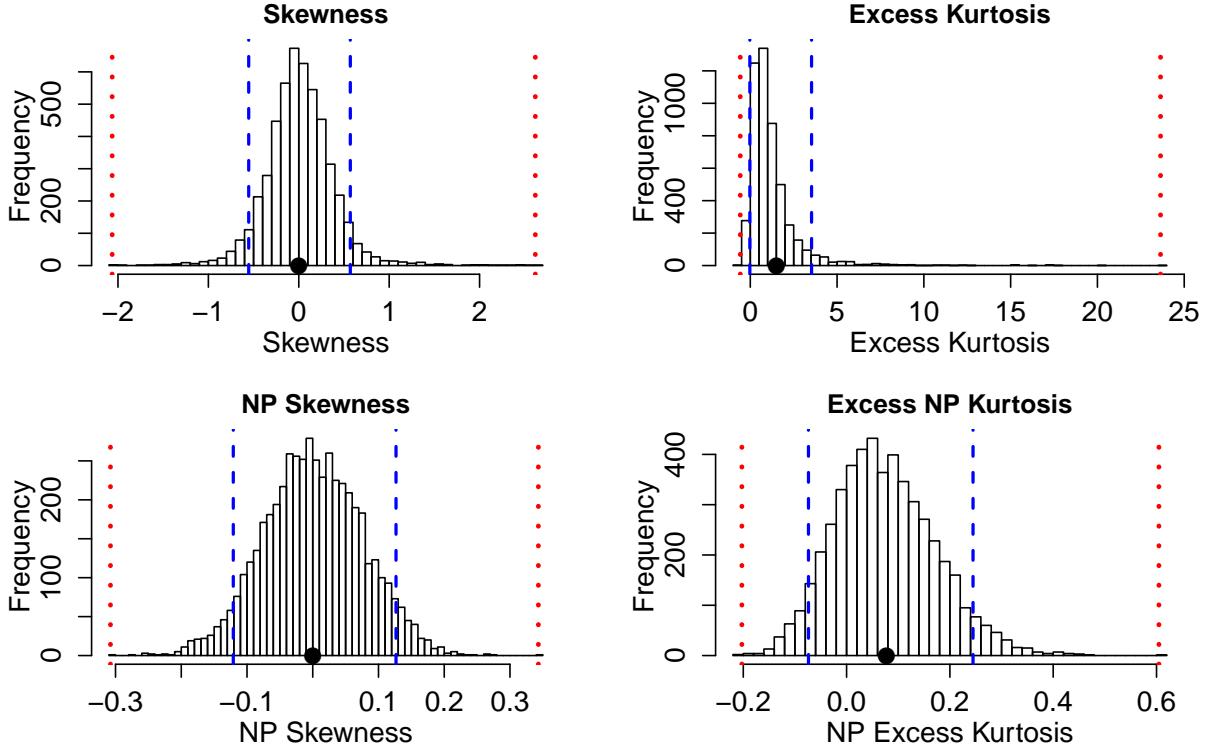


Figure 2: Histograms of skewness and kurtosis estimates using moment-based definitions, compared with the nonparametric measures. Histograms are based on 5000 replicate draws of a sample of 200 independent values from a  $t$  distribution with 8 degrees of freedom. Dotted red vertical lines mark the 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`

ple of size 200 from a  $t$  distribution with 8 degrees of freedom; the true skew is 0, and the true excess kurtosis is 1.5. The distance between the largest and smallest estimates (indicated by the dotted red vertical lines), relative to the distance between the 5th and 95th percentiles, shows the broad extent of extreme values that can occur even with a large sample, especially for kurtosis.

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

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

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

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

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

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

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

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

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

244 Quantile-based estimation of skewness and kurtosis carries the added value that  
245 quantile regression methods may be used to derive these properties of size transitions  
246 as continuous functions of initial size or expected future size. In the examples below, we  
247 use the **qgam** package to fit smooth additive quantile regression models, which have the  
248 flexibility to accommodate non-linear size-dependence in skewness and kurtosis. One  
249 risk of a gam-based approach is that fitted quantiles may be too “wiggly” without con-  
250 straints on their complexity (in the examples below, we specify fitting a spline with  $k = 4$

251 basis functions). For the gam-averse, other quantile regression models may be equally  
252 suitable. For consistency with non-parametric skewness and kurtosis, we similarly use  
253 quantile-based measures of location and scale, and use quantile regression to visualize  
254 these as functions of size. Specifically, following Wan et al. (2014),

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

255 and

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

## 258 1 Case study: Sea fan corals, *Gorgonia ventalina*

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

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

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

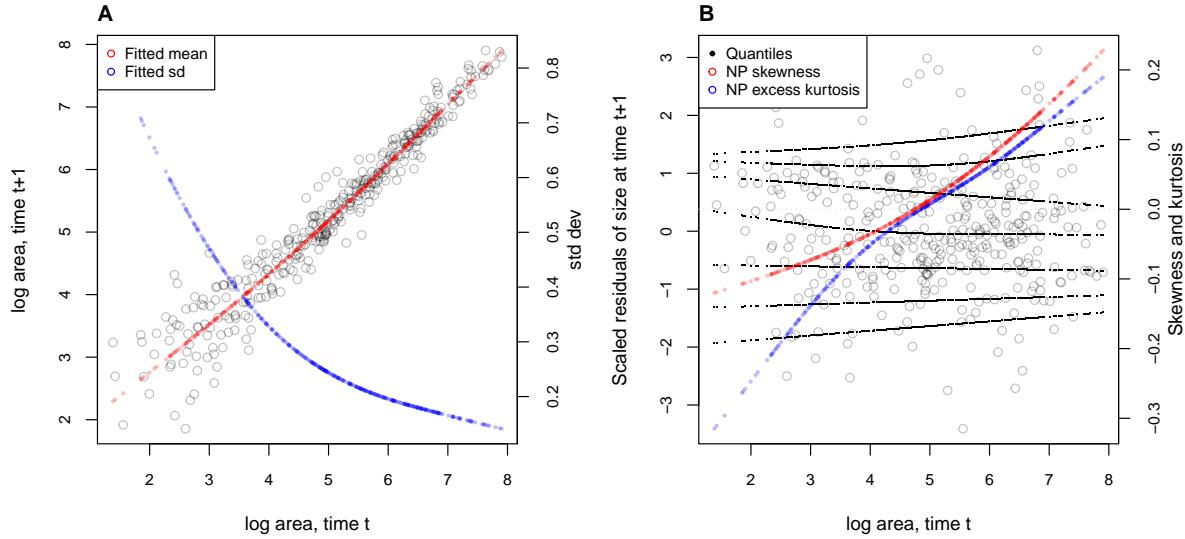


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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_all = predict(fitGAU,type="response");
284 fitted_sd = 1/fitted_all[,2];
285 scaledResids = residuals(fitGAU,type='response')/fitted_sd;

```

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

There are no blatant signs of trouble in the pilot Gaussian model, but quantile regressions on the scaled residuals, and the NP Skewness and Kurtosis metrics derived from them (Eq. 3 and 4), suggest deviations from normality (Fig. 3B). Specifically, skewness switches from negative to positive across the size range, with smaller corals more prone to extreme shrinkage and larger corals more prone to extreme growth. Kurtosis also changes direction over the size distribution, with thinner tails than Gaussian at small sizes and fatter tails at large sizes. The fitted nonparametric moments suggest that the upper and lower tails of size transition probabilities may differ by up to 20%, and the weight of the tails may be 20% greater or less than Gaussian, depending on initial size – not overwhelming deficiencies, but not trivial either. Are these deviations from normal-

300 ity severe enough to warrant a second, non-Gaussian iteration of growth modeling? To  
301 answer that question, we simulated data from the fitted Gaussian model and examined  
302 whether key properties of the simulated data are consistent with those of the real data –  
303 this is the ultimate litmus test for a growth model's adequacy and should be a standard  
304 element of IPM construction, in our opinion. If the simulated data are not consistent  
305 with the real data, it is time to choose a better distribution (Fig. 1). In this case, most  
306 of 100 Gaussian model simulations are out of line with the skew at smallest sizes and  
307 largest sizes, and the excess kurtosis observed at moderately large sizes (Fig. 4 CD). For  
308 at least some parts of the size distribution, a non-Gaussian model would better capture  
309 size transitions.

310 We sought a distribution that could accommodate the observed changes in the sign  
311 of skewness and excess kurtosis. We chose the sinh-arcsinh (SHASH) distribution, a four-  
312 parameter distribution that, conveniently, is included in **mgcv**'s **gam()** function: **SPE: I**  
313 **made the SHASH model consistent with the Gaussian for location and scale, removing**  
314 **the specification of k.**

```
315     fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0), # <- location  
316                     ~ s(logarea.t0), # <- log-scale  
317                     ~ s(logarea.t0,k=4), # <- skewness  
318                     ~ s(logarea.t0,k=4)), # <- log-kurtosis  
319                     data = XH, gamma = 1.4, family = shash, optimizer = "efs")
```

320 The fitted model's mean and variance are nearly identical to the Gaussian (Fig. 4AB),  
321 and the fitted trends in skewness and kurtosis are much less “wiggly” than the estimate  
322 from the data (Fig. 4CD). Nonetheless, data simulated from the SHASH model are more  
323 consistent with the real data, with more SHASH data sets matching or exceeding the  
324 largest skewness and kurtosis values observed (Fig. 4CD). If one cares to quantify the  
325 difference between models, the SHASH model is clearly favored by AIC ( $\Delta AIC = 5.45$ )  
326 despite having twice as many parameters to fit.

327 What, then, have we gained by fitting a better growth model? Fig. 5A compares  
328 the predicted distributions of subsequent size in the fitted model and Gaussian pilot  
329 models, for the median size of a new recruit (leftmost pair of curves), the median ini-  
330 tial size (central curves), and the 95th percentile of initial size in the data (rightmost  
331 curves). The differences are small, and most pronounced for the smallest size, where re-  
332 cruits are predicted to grow slightly larger under the SHASH model than the Gaussian  
333 model. The direction of this difference was surprising, because the SHASH has negative  
334 skew at small sizes in the data. However, the SHASH model also gives a better predic-

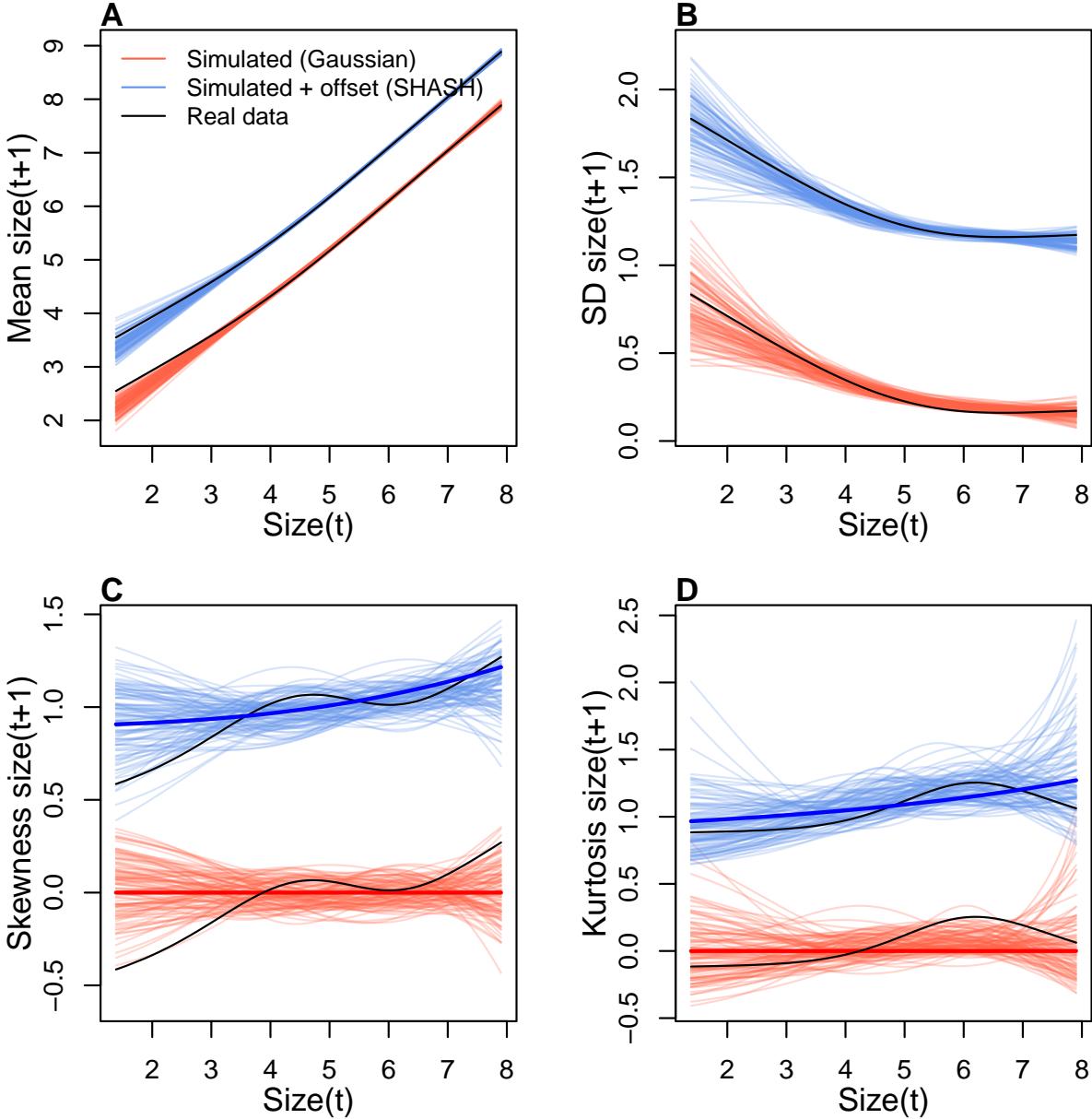


Figure 4: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Note that plotted values for the SHASH are offset by one unit to allow comparisons. In the skewness and kurtosis panels, the darker solid curves show the values for the fitted growth models. Figure made by script AkumalCorals\_qgam.R.

335 tion of mean growth at small sizes than the Gaussian model. At intermediate sizes the  
 336 predictions are nearly identical; at large sizes the SHASH has slightly lower standard  
 337 deviation, but fatter tails (excess kurtosis). Fig. 5B shows the predicted steady-state size

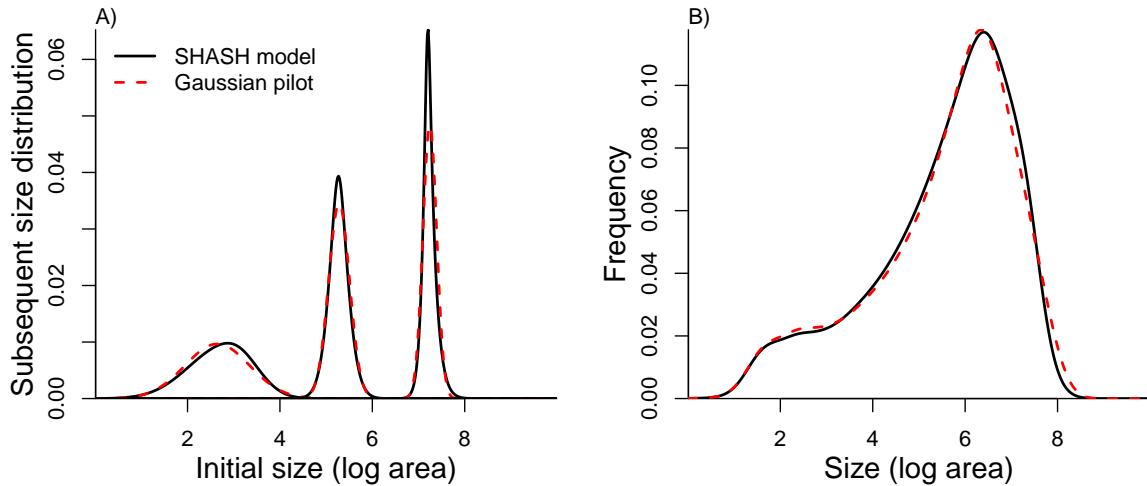


Figure 5: Comparisons between the fitted SHASH growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year  $t + 1$  for three different values of size in year  $t$ . The leftmost pair of curves are 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 measured sizes of known new recruits ( $n = 9$ ). Figure made by script AkumalCoralsIPMs.R.

338 distributions resulting from a constant unit input of recruits. Again, the differences are  
 339 very subtle. Finally, the Gaussian and SHASH growth models predict very similar mean  
 340 life span (17.7 and 17.9 years, respectively).

341 From these outputs, there is little evidence that improved modeling of coral growth  
 342 meaningfully improved biological inferences from the IPM. One could argue that it was  
 343 not worth the trouble. But it was almost no trouble at all, because building an IPM with  
 344 a SHASH model is almost line-by-line identical to using a Gaussian model. And without  
 345 fitting the SHASH model, it is not possible to know whether or not it would have made  
 346 a difference.

347 In this case study we used `gam` to fit both the Gaussian and SHASH models because  
 348 that obviated model selection on functions for mean, variance, and higher moments.  
 349 However, `gam` should be used with caution. Nonparametric regression models notori-  
 350 ously “wag their tails” because the ends of the fitted curve can be pulled close to the

outermost data points. This is especially problematic for growth modeling, because data are typically sparse near the bounds of the size distribution. To minimize the risk of overfitting we specified the number of “knots” ( $k=4$ ) and used  $\text{gamma}=1.4$  to overweight model degrees of freedom as suggested by Gu (2013, sec. 3.2). But it is always important to plot the fitted splines and make sure they do not wag unrealistically. If they do, parametric regression may be a better choice.

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

The next case study, focusing on the tree cholla cactus *Cylindriopuntia imbricata* at the Sevilleta Long-Term Ecological Research site in central New Mexico, adds a new feature on top of the simple size-dependent regressions in the previous study: random effects associated with temporal (year) and spatial (plot) environmental heterogeneity. This long-term study of cactus demography was initiated in 2004 and different subsets of the data have been analyzed in various IPM studies, all using Gaussian growth kernels (Compagnoni et al., 2016; Czachura and Miller, 2020; Elderd and Miller, 2016; Miller et al., 2009; Ohm and Miller, 2014). In fact, (Elderd and Miller, 2016) presented a Gaussian growth model fit to the cactus data as an example of a well fit growth function, based on a marginal distribution of residuals that appeared approximately Gaussian and posterior predictive checks (PPCs) of a Bayesian model that suggested consistency between the real data and data simulated from the fitted model (Fig. 4 in (Elderd and Miller, 2016)).

While PPCs and the associated “Bayesian P-value” are popular diagnostic tools, they are often considered to be too conservative (Conn et al., 2018; Zhang, 2014), failing to reject marginally bad models even though they are very effective in rejecting models that are terrible. The choice of discrepancy function (the statistic used to compare real and simulated data) can also be limiting: in our previous work, we used a discrepancy function focused on variance (the sum of the squared residuals), so we had a built-in blind-spot for mismatches in higher moments. In the clarity of hindsight, the PPC gave a false sense of security; the Gaussian was a poor choice all along.

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

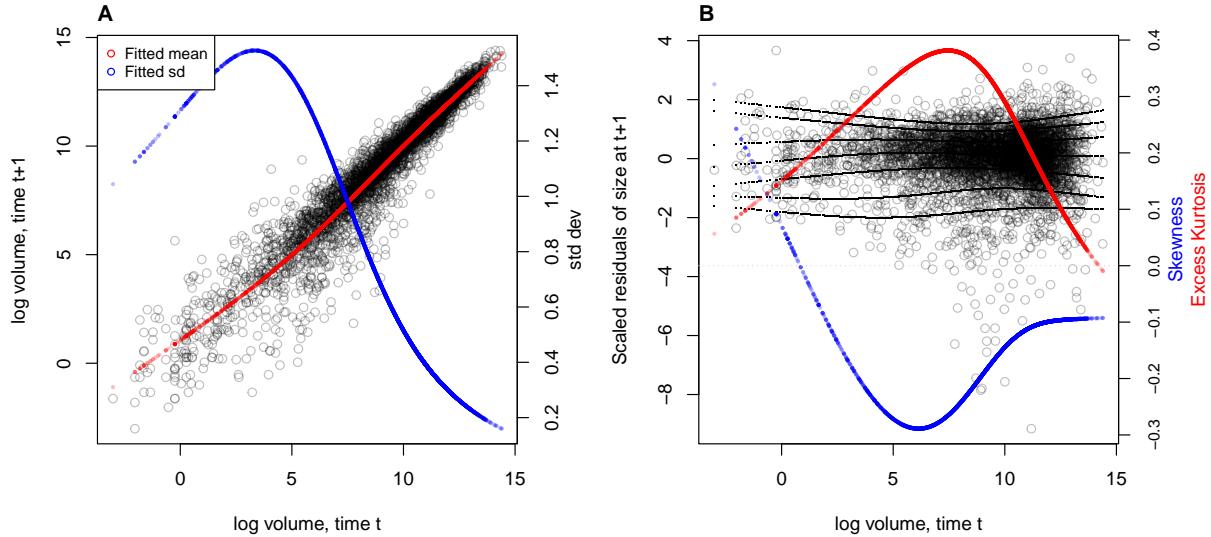


Figure 6: **A**, Size transition data for tree cholla cacti, *Cylindropuntia imbricata*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess 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`.

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

391 To better capture size transitions, we need a distribution with negative skew and  
 392 positive excess kurtosis, but both of which may be negligible at some sizes. We first tried  
 393 Johnson's  $S_U$  and then the skewed  $t$  distributions, both of which are limited to positive  
 394 excess kurtosis. Both distributions provided some improvement over the Gaussian, but  
 395 were not happy with the fit of either. Iterating through the workflow (Fig. 1), we ar-  
 396 rived, again, at the SHASH distribution, which is more flexible than either the JSU or  
 397 skewed  $t$ , capable of capturing a greater range of kurtosis for a given amount of skew,  
 398 and vice versa (Steve's NPSkewKurtosisRanges.pdf). Furthermore, fitting the SHASH  
 399 as a generalized additive model with `mgcv` allowed for flexible, non-monotonic size-  
 400 dependence in skewness and kurtosis without the need for model selection on specific

401 size-dependent functions; through iterations of trial and error, we found this flexibility  
 402 was necessary to generate simulated data that compared favorably to the real data. The  
 403 other distributions that we tried are not available as `mgcv` families, so we fit these with  
 404 custom maximum likelihood functions, an approach we illustrate in the next case study.  
 405 The final growth model was similar to the SHASH gam in the coral case study, but  
 406 with random intercepts for the location parameter, representing spatial and temporal  
 407 heterogeneity:

```

408 fit_shash <- gam(list(logvol_t1 ~ s(logvol_t,k=4) +
409   s(plot,bs="re") + s(year_t,bs="re"), # location
410   ~ s(logvol_t,k=4), # log-scale
411   ~ s(logvol_t,k=4), # skewness
412   ~ s(logvol_t,k=4)), # log-kurtosis
413   data = CYIM_grow,
414   family = shash,
415   optimizer = "efs")
  
```

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

425 We explored how improved growth modeling influenced IPM results, leveraging  
 426 the plot and year structure of the study design to quantify spatial and temporal vari-  
 427 ance in fitness. We used the fitted random effects from the vital rate models to estimate  
 428 the asymptotic growth rate for each year ( $\lambda_t$ ), centered on the average plot, and for  
 429 each plot ( $\lambda_p$ ), centered on the average year. This allowed us to quantify demographic  
 430 variance associated with temporal and spatial heterogeneity. We found that the Gaus-  
 431 sian growth model tended to over-estimate  $\lambda_t$ , particularly in the harshest years (Fig.  
 432 8A), and thus under-estimated temporal variance in fitness ( $Var(\lambda_{t(Gaussian)}) = 0.0018$ ,  
 433  $Var(\lambda_{t(SHASH)}) = 0.0023$ ). The opposite was true for plot-to-plot variation (Fig. 8B),  
 434 where the Gaussian model under-estimated  $\lambda_p$  and over-estimated spatial variance in  
 435 fitness ( $Var(\lambda_{p(Gaussian)}) = 0.00015$ ,  $Var(\lambda_{p(SHASH)}) = 0.000088$ ). Across both growth

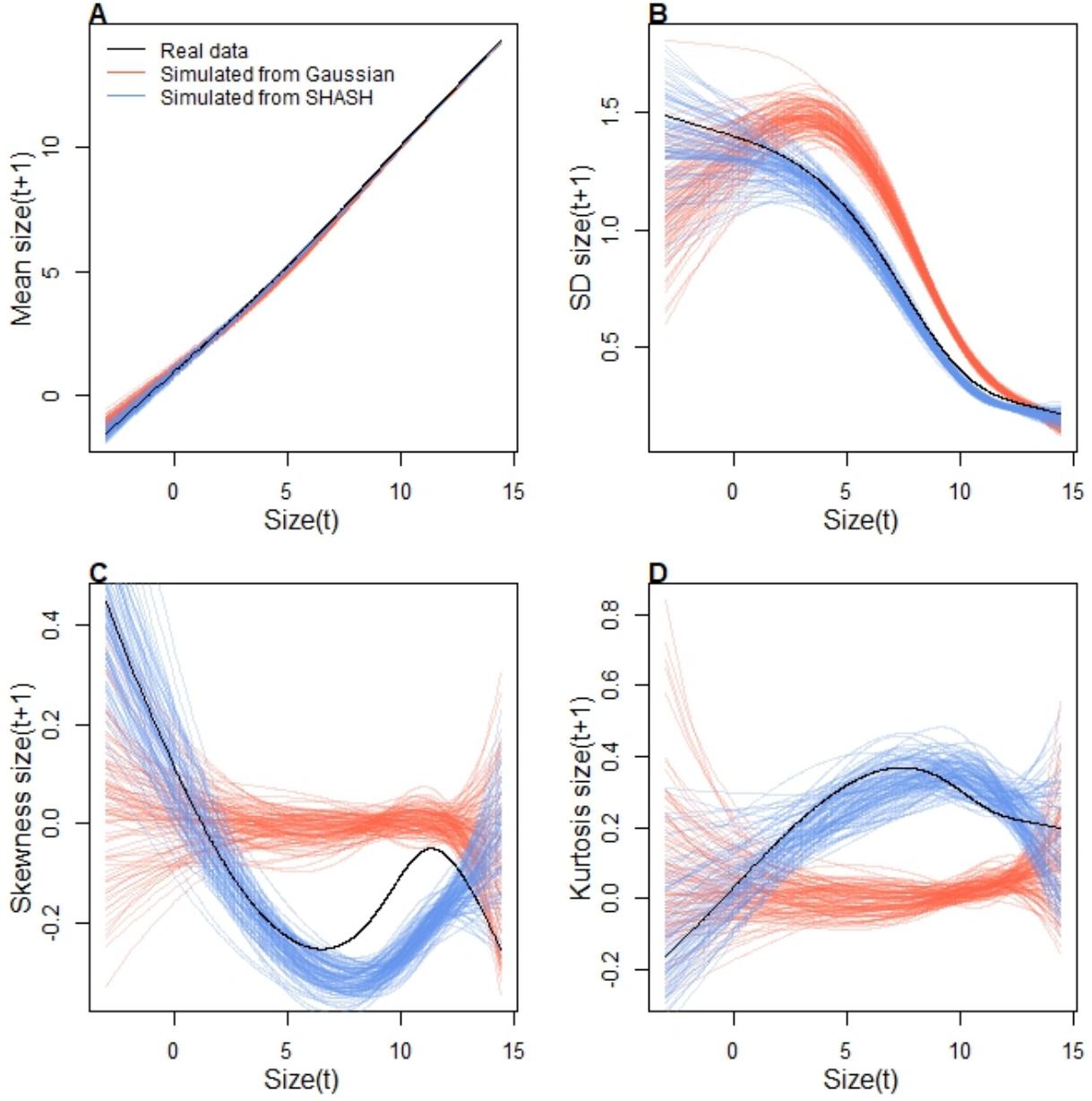


Figure 7: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

models, fluctuations in fitness were stronger through time than across space. The difference in temporal variance would suggest that Gaussian growth modeling would lead to over-estimation of the stochastic growth rate  $\lambda_S$ , since temporal variance has a negative influence on  $\lambda_S$ . However, this was not the case: stochastic IPMs based on Gaussian and SHASH growth models had nearly identical stochastic growth rates

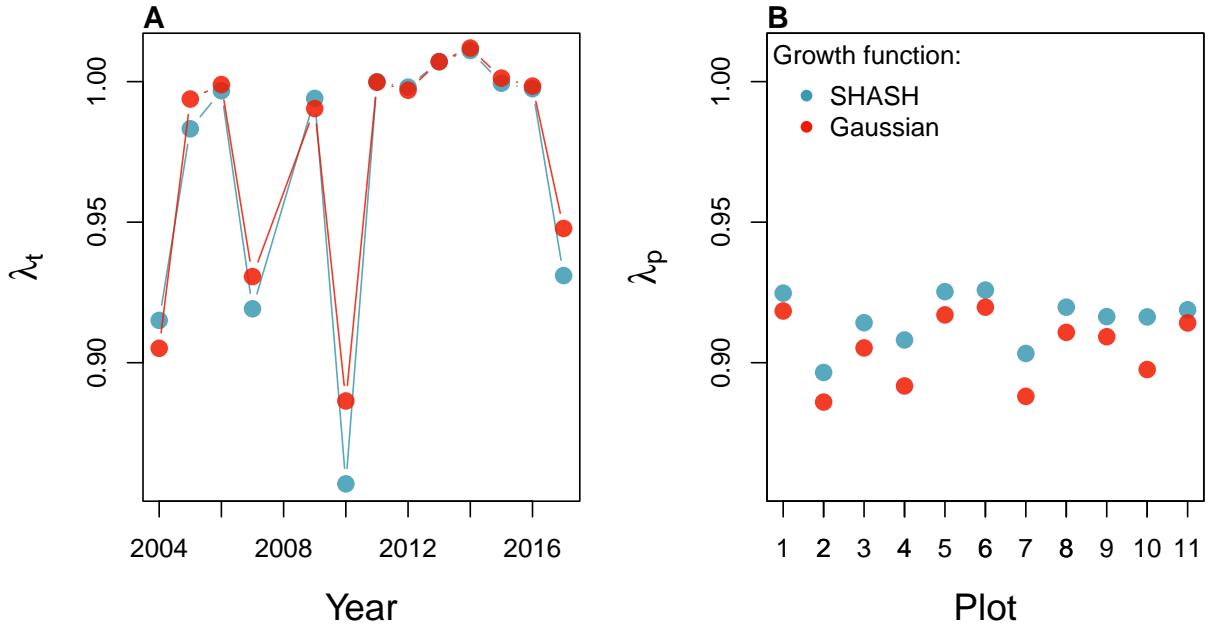


Figure 8: Temporal (A) and spatial (B) heterogeneity in fitness for the tree cholla cactus (*Cylindropuntia imbricata*) predicted by IPMs using Gaussian or SHASH growth models. Figure made by script `cactus_growth_modeling_qgam.R`.

(441)  $(\lambda_S(\text{Gaussian}) = 0.9906, \lambda_S(\text{Gaussian}) = 0.9909)$ . This is likely because temporal fluctu-  
 (442) ations in vital rates, which is where the SHASH growth model would make a difference,  
 (443) have a weaker influence on  $\lambda_S$  than the temporal fluctuations in size structure that they  
 (444) generate (Compagnoni et al., 2016; Ellis and Crone, 2013). Thus, depending on the target  
 (445) of one's analysis, modeling non-Gaussian size transitions with a Gaussian growth model  
 (446) could bias results in either direction, or make no difference at all.

### 447 3 Case study: creosotebush, *Larrea tridentata*

448 Our next case study comes from our studies of the woody shrub creosotebush (*Larrea tri-  
 449 dentata*) at the Sevilleta Long-Term Ecological Research (LTER) site in central New Mex-  
 450 ico, US. At this site as elsewhere in the Southwest US, creosotebush is encroaching into  
 451 desert grassland habitats. The data described here were collected along transects span-  
 452 ning grass-shrub ecotones to understand patterns of density dependence in creosotebush  
 453 demography. Specifically, we asked whether fitness is maximized approaching zero den-  
 454 sity at the leading edge of the expansion front (consistent with ‘pulled’ expansion), or  
 455 whether there is a demographic advantage for shrubs at higher density due to positive  
 456 feedbacks expected for ecosystem engineers (leading to ‘pushed’ expansion). Our pub-

457 lished study (Drees et al., 2023) used a spatial integral projection model (SIPM) to predict  
458 the speed of shrub encroachment, assuming normally-distributed size transitions. Here  
459 we step through our suggested workflow to ask whether a non-Gaussian model would  
460 have been more faithful to the data, and how such an improvement would influence  
461 predictions for the speed of encroachment. We use this case study to illustrate several  
462 new elements and challenges, including modeling skewness and kurtosis as functions  
463 of expected future size (instead of initial size) and using distributions that are not cur-  
464 rently available as **mgcv** families. In fact, to diversify our use of software and illustrate  
465 alternatives, we do not use gam's for any element of this case study.

466 Growth data come from 522 shrubs censused longitudinally over four years (2013-  
467 2017). Census individuals occurred along 12 replicate transects (200 to 600 m in length)  
468 that spanned gradients of shrub density along shrub-grass ecotones. Size was measured  
469 as volume of an elliptical cone based on height and width measurements; the size vari-  
470 able of the IPM was the natural logarithm of volume ( $cm^3$ ). For each census individual,  
471 we recorded the size and density of all conspecifics within the five-meter transect “win-  
472 dow” in which it occurred, and took the sum of all sizes within the window as a measure  
473 of local density. The data are available in Ochocki et al. (2023).

474 As an initial Gaussian approach, we first fit a set of candidate generalized linear  
475 mixed models, including transect as a random effect, that represented competing hy-  
476 potheses for how size, density, and their interaction influence growth. Specifically, we fit  
477 five candidate Gaussian models that included fixed effects of initial size only (model 1),  
478 size and density (model 2), and size, density, and their interaction (model 3), allowing  
479 for shrubs of different sizes to have different growth responses to local density. Models  
480 4 and 5 mirrored models 2 and 3 but included second-order terms for density, allowing  
481 for the possibility of non-monotonic density dependence. As in (Drees et al., 2023) we  
482 pooled data across three transition years. Initial AIC rankings of these pilot models fa-  
483 vor model 4 slightly over model 5 ( $\Delta AIC = 0.8$ ) and significantly over all other models  
484 ( $\Delta AIC > 2$ ). However, these models were fit assuming constant variance, and inspection  
485 of the residuals of the best model indicate this is not a safe assumption.

486 Unlike our previous case studies, here we have multiple fixed effects that may influ-  
487 ence the variance of future size. In cases such as this, we recommend modeling variance  
488 as a function of expected future size rather than initial size, as we did with the corals  
489 and cacti. The expected (or “fitted”) values reflect the combined influence of all fixed  
490 and random effects, and therefore implicitly account for multiple sources of variation in  
491 the variance. While there are several convenient software packages for simultaneously  
492 modeling Gaussian mean and variance as functions of independent variables (**mgcv** for

493 additive models as we saw above, **nmle** for linear models), **modeling variance as a function**  
 494 **of the mean is trickier because they cannot easily be fit simultaneously**<sup>1</sup>. Here we  
 495 us an iterative re-weighting approach – which is not elegant, but it works. For Gaus-  
 496 sian models, weights  $w_i$  can be used to indicate that the observations  $y_i$  vary in their  
 497 dispersion around the mean. In general, the iterative steps are:

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:

$$y_i = \mu_i + \epsilon_i$$

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

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

505 The resulting Gaussian growth model predicts strong initial size-dependence and  
 506 weak and slightly nonlinear (but monotonic) negative density dependence (Fig. 9A).  
 507 The model accounts for non-constant variance through the fitted weights, which indicate  
 508 greater dispersion for smaller values of expected size ( $\beta_1 = -0.21$ ; Fig. 9B). Quantiles of  
 509 the standardized residuals indicate weak negative skew (difference in tail size is 1–2%  
 510 of their total) and positive excess kurtosis, especially at smaller expected sizes (tails are

---

<sup>1</sup>After I wrote this I discovered that **nlme** can fit residual variance as a function of **fitted(.)**.

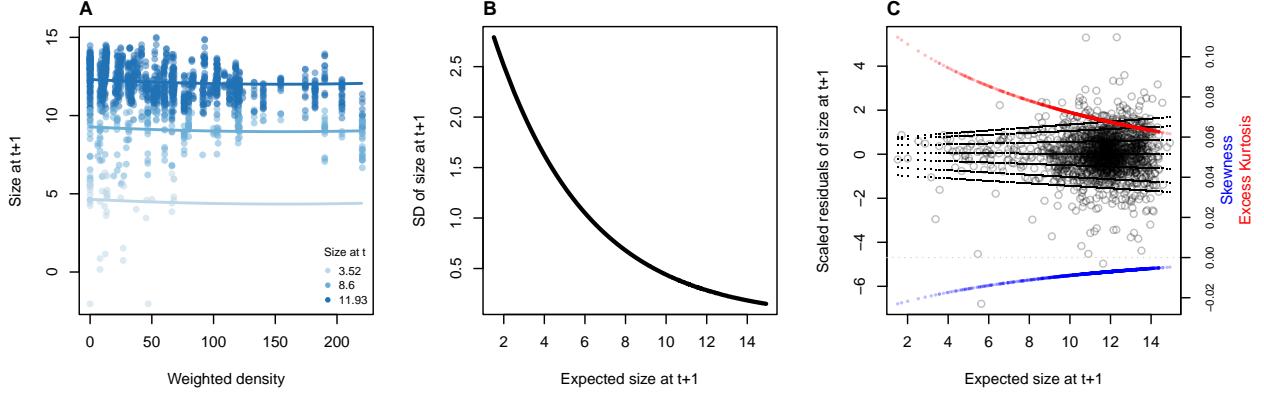


Figure 9: **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume ( $cm^3$ ). **B**, Standard deviation of size at time  $t + 1$  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`.

511 6–10% fatter than Gaussian) (Fig. 9C).<sup>2</sup> As a candidate for improvement, we turned to  
 512 the Johnson’s  $S_U$  (JSU) distribution, a four-parameter, leptokurtic distribution capable  
 513 of skew in either direction. We used a parameterization of the JSU for which location  
 514 parameter  $\mu$  is the mean and scale parameter  $\sigma$  is the standard deviation (Rigby et al.,  
 515 2019).

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

```
524 ## log_volume_t1 are the size obervations
525 ## GAU_fitted are the expected values of the best Gaussian model
526 ## pars is a vector of free parameters to be estimated
```

---

<sup>2</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.

```

527 JSULogLik=function(pars){
528   dJSU(x=log_volume_t1,
529   mu=GAU_fitted,
530   sigma=exp(GAU_sd_coef[1]+GAU_sd_coef[2]*GAU_fitted),
531   nu = pars[1]+pars[2]*GAU_fitted,
532   tau = exp(pars[3]+pars[4]*GAU_fitted), log=TRUE)
533 }
```

534 The mean of the JSU is set to that of the best Gaussian model (GAU\_fitted) and the  
535 standard deviation is a function of the mean according to the coefficients (GAU\_sd\_coef)  
536 estimated through iterative re-weighting. Based on diagnostics of the standardized resid-  
537 uals (Fig. 9), JSU parameters that control skewness and kurtosis are defined as linear  
538 functions of the mean, and it is these coefficients that are estimated by maximum like-  
539 lihood. Here we are relying on the robustness of Gaussian linear models to deviations  
540 from normality . If one is skeptical of this approach, it is possible, as an alternative,  
541 to simultaneously re-fit all parameters of the JSU in a maximum likelihood framework.  
542 However, incorporating random effects into a custom likelihood model is non-trivial (we  
543 provide guidance on one way to do this, using the “shrinkage” approach, in Appendix  
544 XX). Therefore a key advantage of the hybrid approach is convenient retention of the  
545 fitted random effects and associated variance components, which get shuttled from the  
546 Gaussian model into the non-Gaussian model without any fuss (it was critical that we  
547 used a parameterization of the JSU for which `mu` is the mean and `sigma` is the standard  
548 deviation). And, if this approach does not “work” (i.e., deviations from normality bi-  
549 ased the fitted values of the Gaussian model) one would quickly find out through the  
550 simulation step of the workflow. In this case, the hybrid JSU model performed well,  
551 generating simulated data that aligned with the real data better than the best Gaussian  
552 model, particularly in **standard deviation**<sup>3</sup> and kurtosis (Fig. 10). Note that in Fig. 10  
553 we are plotting moments of the future size distribution with respect to initial size; this  
554 distribution is also conditional on density but initial size is by far the stronger predictor  
555 of future size, so we chose this visualization.

556 The improvement of the JSU over the Gaussian growth model, while visually satis-  
557 fying, had virtually no influence on SIPM results. Models using Gaussian or JSU growth  
558 kernels had nearly identical, monotonic decreases in  $\lambda$  with increasing local density, and  
559 nearly identical wave velocities (Fig. 11). This species has very low mortality risk once  
560 established (mean remaining life expectancy of a median-sized shrub is 24,408 years)

---

<sup>3</sup>I am a little mystified as to why the JSU is so much better. It is literally the same SD in both distributions.

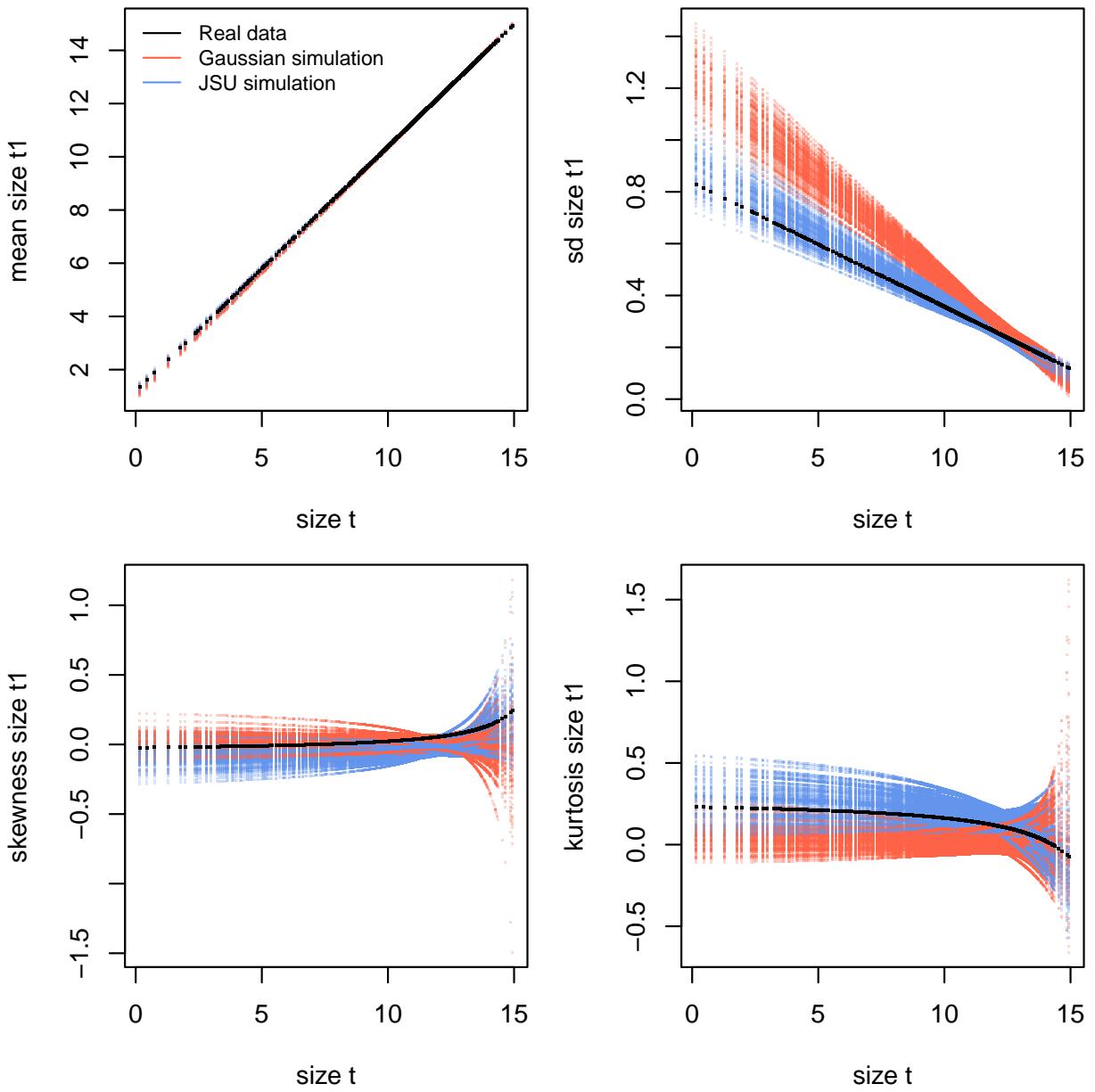


Figure 10: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `creosote_growth_modeling_qgam.R`.

and its population growth and wave expansion are limited by very low seedling recruitment ((Drees et al., 2023)). Weak size-dependence in survival likely explains why the improvement in growth modeling had little influence on SIPM predictions.

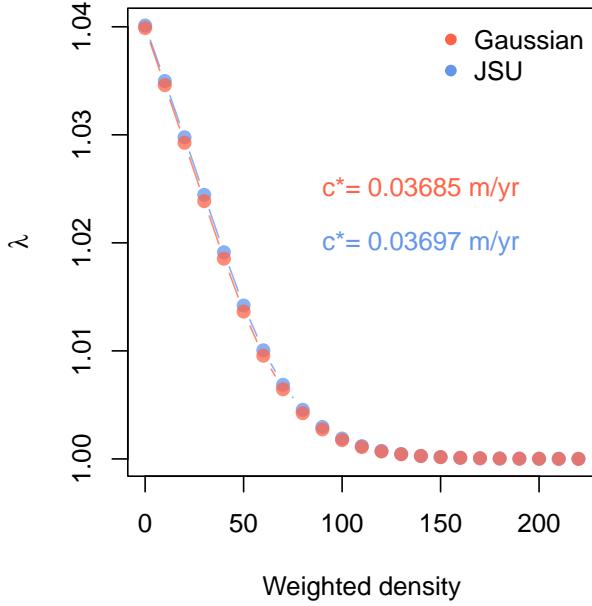


Figure 11: Density dependence in fitness ( $\lambda$ ) and asymptotic velocity of the creosote encroachment wave ( $c^*$ ) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ( $\log(cm^3)$ ) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

## 564 4 Case study: lady orchid, *Orchis purpurea*

565 Our final case study examines selection on life history strategies in the lady orchid *Or-  
566 chis purpurea*. In a prior study, Miller et al. 2012 contrasted the growth trajectories from  
567 year  $t$  to  $t + 1$  for plants that did or did not flower in year  $t$ , as a way to quantify costs  
568 of reproduction. The different growth kernels were then used in an IPM to quantify  
569 evolutionarily stable life history strategies: the optimal flowering size that balances ben-  
570 efits of flowering at larger sizes against the risk of dying before reaching those sizes.  
571 The original study assumed a Gaussian distribution of size transitions and allowed for  
572 non-constant variance with respect to initial size. Here we re-visit that analysis applying  
573 our growth modeling workflow to derive improved growth kernels for flowering and  
574 non-flowering orchids.

575 The data, originated by Dr. Hans Jacquemyn and used here with permission, come  
576 from 368 plants in a Belgian population that was censused annually from 2003 through  
577 2011 (for this reanalysis we are using data only from the “light” habitat). Size was mea-  
578 sured as leaf area ( $cm^3$ ) summed over all leaves, and we analyzed the natural logarithm  
579 of total leaf area as the size variable of the IPM.

580 As a pilot Gaussian approach, we fit six candidate models in which the mean was  
 581 a function of initial size only, additive effects of initial size and flowering status, and  
 582 interaction between size and flowering, and the standard deviation was a function of  
 583 size only (models 1-3) or size and flowering status (models 4-6). All models included a  
 584 random intercept for year. As another variation on software and an alternative to two-  
 585 step fitting or iterative re-weighting, here we use `nmle::lme()`, which can simultaneously  
 586 fit linear predictors for mean and variance. For example, model 1 was:

```

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

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

597 As improvements, we explored the skewed  $t$  and Johnson's SU distributions, both  
 598 leptokurtic distributions with flexible skewness. We were happier with the skewed  $t$ ,  
 599 which we fit in a similar way as we fit the JSU to the creosote data, setting the mean  
 600 and standard deviation to the Gaussian fits and estimating free parameters controlling  
 601 skewness and kurtosis:

```

602 ## log_area_t1 and log_area_t are the size obervations
603 ## flowering indicates reproductive status at time t (0 or 1)
604 ## GAU_fitted and GAU_sd are mean and standard deviation from lme
605 ## pars is a vector of free parameters to be estimated
606 SSTLogLik=function(pars){
607   dSST(x=log_area_t1,
608         mu=GAU_fitted,
609         sigma=GAU_sd,
610         nu = exp(pars[1] + pars[2]*log_area_t + pars[3]*as.logical(flowering) + pars[4])
611         tau = exp(pars[5] + pars[6]*log_area_t + pars[7]*as.logical(flowering) + pars[8]
612         log=TRUE)
613 }
  
```

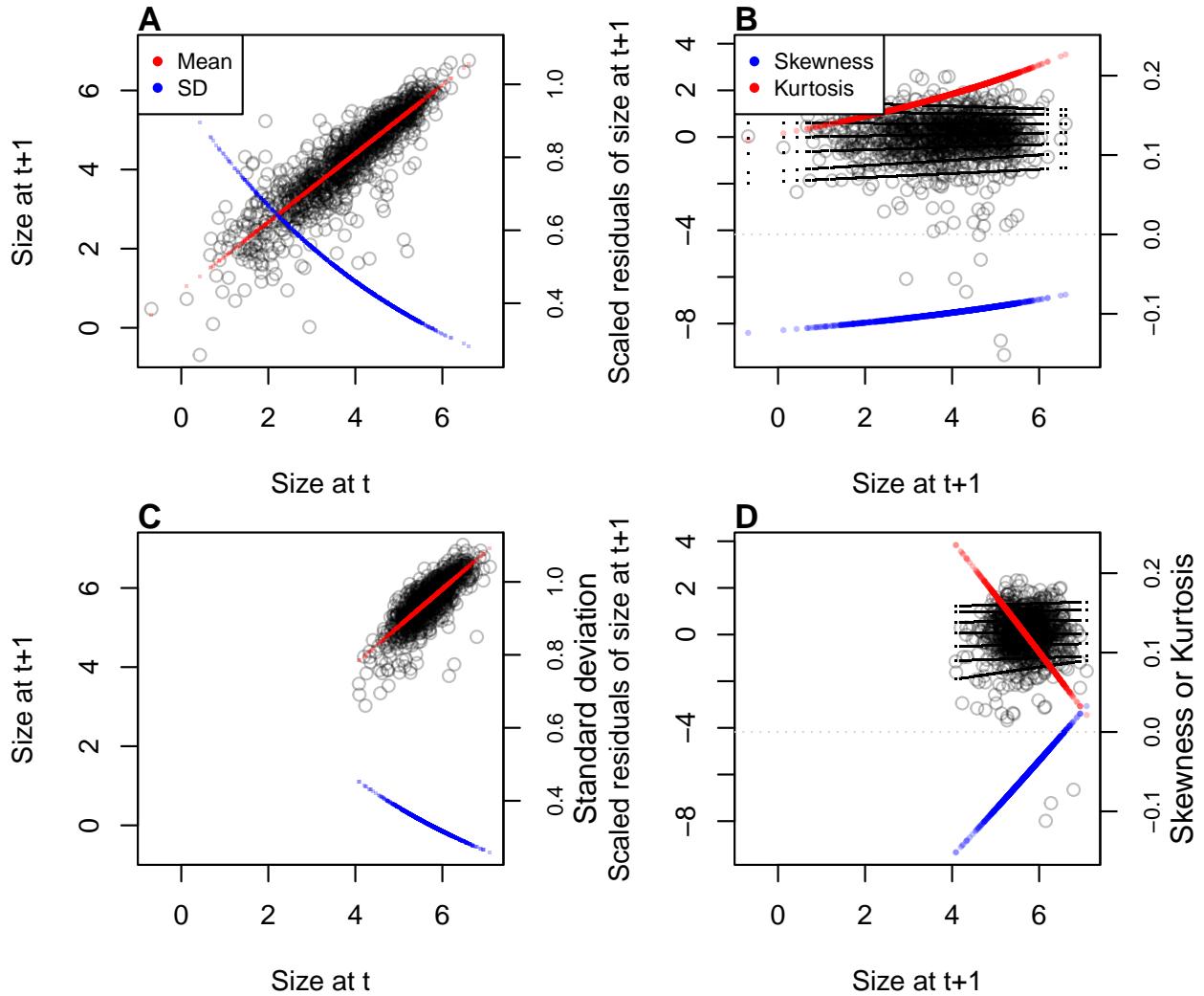


Figure 12:

614 `gamlss.dist:dsST` is a parameterization of the skewed  $t$  in which `mu` and `sigma` are the  
 615 mean and standard deviation, respectively. Based on diagnostics of the standardized  
 616 residuals (Fig. 12) we allowed `nu` and `tau` to vary by size and differ between flowering  
 617 and non-flowering plants (note that the `tau` parameter uses a  $\log(x - 2)$  link function).  
 618 Size transition data simulated from this model corresponded favorably to the real data,  
 619 much better than the pilot Gaussian model, including improvements in the **standard**  
 620 **deviation**<sup>4</sup>, skewness, and kurtosis of future size (Fig. 13).

621 Finally, we used the improved growth model to revisit key results of the original  
 622 study. Miller et al. (2012) used the orchid IPM to estimate the evolutionarily stable strat-  
 623 egy (ESS) as the mean size at flowering that maximizes lifetime reproductive success

<sup>4</sup>Again, the improvement here is surprising to me and I am unsure what to say about it.

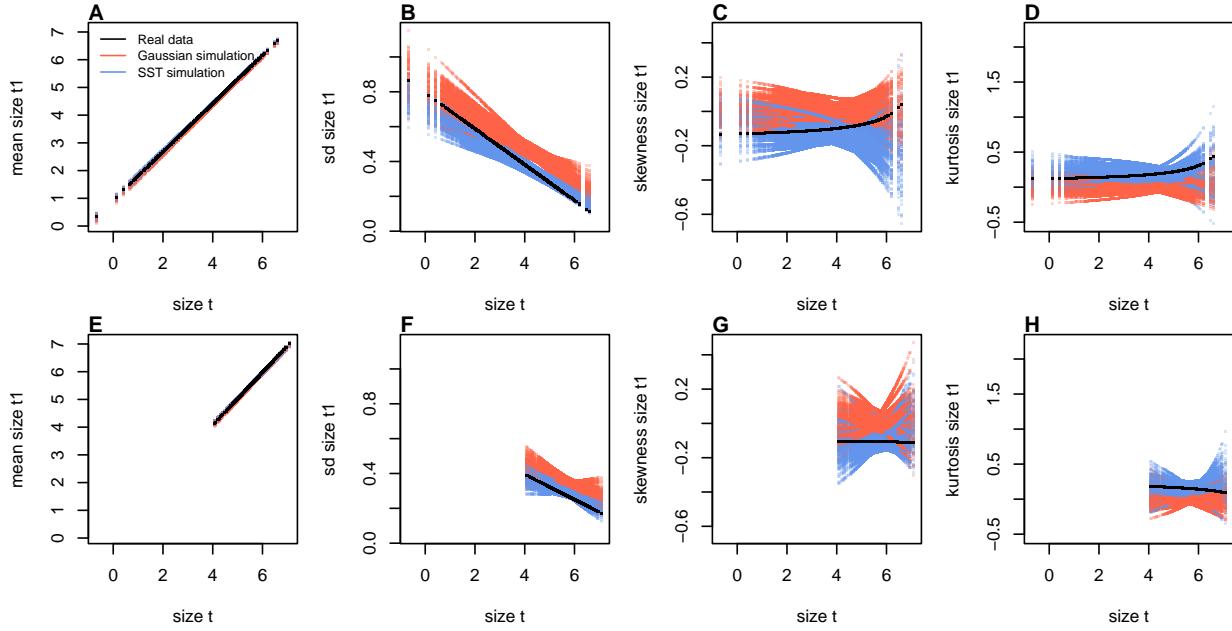


Figure 13: Comparisons between real orchid data and data simulated from Gaussian and skewed  $t$  growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (A-D) shows plants that were vegetative at the start of the transition year and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

624 ( $R_0$ ), given the constraint that flowering when small reduces growth and thus elevates  
 625 mortality risk. Repeating that analysis here, we found that improved growth modeling  
 626 has virtually no influence on predictions for optimal life history strategies (Fig. 14). ESS  
 627 flowering sizes were nearly identical between IPMs with Gaussian vs skewed  $t$  growth  
 628 models, and both aligned well with the observed mean flowering size (dashed vertical  
 629 line in Fig. 14A). Extending beyond the original study, we also explored expected re-  
 630 maining lifespan for different ages and sizes (R package **Rage** (Jones et al., 2022)). Gaus-  
 631 sian and skewed  $t$  growth models predicted nearly identical mean remaining lifespans  
 632 across the stage and size distribution (Fig. 14B). However, the skewed  $t$  model predicted  
 633 consistently greater variance in remaining lifespan, nearly 10% greater at some sizes.<sup>5</sup>  
 634 Thus, as we have seen in other case studies, the practical consequences of improved  
 635 growth modeling depend on what one aims to learn from the IPM.

<sup>5</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.*

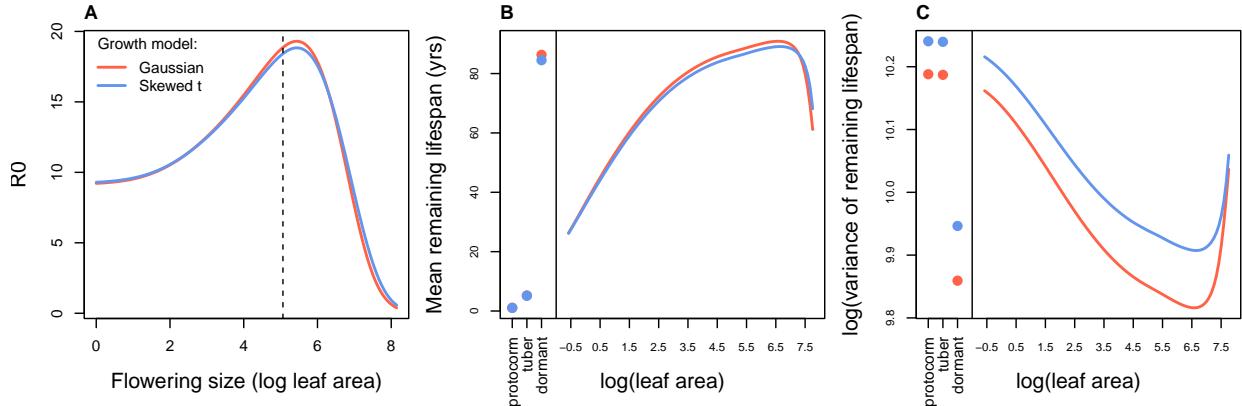


Figure 14: Orchid life history results from IPMs using Gaussian or skewed *t* growth models. **A**, Lifetime reproductive success ( $R_0$ ) as a function of mean size of flowering. Dashed vertical line shows the observed mean flowering size. **B-C**, Mean and variance of remaining lifespan as a function 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.

## 636 5 Discussion

637 Much of the appeal of integral projection models has stemmed from their embrace of  
 638 continuous size structure through reliance on regression-based approaches, and the po-  
 639 tentially complex fixed- and random-effect structures that these approaches allow. Using  
 640 familiar statistical tools and with relatively few parameters to estimate, IPM users can  
 641 incorporate important sources of variation in demography and interrogate their influ-  
 642 ence on ecological and evolutionary dynamics. With this opportunity comes the burden  
 643 of getting it right: IPMs are good models of the populations they are intended repre-  
 644 sent only insofar as the statistical models provide good fits to the underlying data. The  
 645 growth sub-model is the trickiest part of “getting it right” because it defines a distri-  
 646 bution of future size conditional on current size. Distributions have many properties –  
 647 “moments” – and a good growth model should recapitulate the properties of real size  
 648 transitions. The default assumption of normally distributed size transitions, employed  
 649 overwhelmingly across 20+ years of IPM studies, is an arbitrary historical precedent.  
 650 In four case studies (chosen simply because we had the data at our fingertips) and,  
 651 we suspect, more broadly, skewness and excess kurtosis were common features of size  
 652 transition data: shrinking was more common than growing, and large changes in size  
 653 were more common than a Gaussian model would predict. Our most important mes-

654 sage is that the standard assumption of normally-distributed size transitions should be  
655 abandoned and a more inquisitive process of growth modeling should take its place.

656 We have attempted to lay out a general workflow for what that process should look  
657 like, guided by visual diagnostics of standardized residuals. One implication of relying  
658 on visual diagnostics is that goodness of fit is in the eye of the beholder. This approach  
659 can empower IPM users to make informed choices, but it is not very prescriptive: we  
660 have not suggested any hard rules for when one or another distribution should be used,  
661 only that a good growth model should generate data that look like the real thing. Al-  
662 ternatively, model selection could be used to identify best-fitting growth distributions  
663 and best-fitting functions for higher moments. However, model selection among growth  
664 distributions with 3-5 parameters, each of which may be functions of state variables or  
665 fitted values, can quickly explode in complexity, and we are not convinced it is worth  
666 the trouble. It is possible to find a good growth model without worrying about which  
667 one is “best”.

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

684 Some issues to be discussed.

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

689 **Acknowledgements**

690 This research was supported by US NSF grants DEB-1933497 to SPE and DEB-1754468,  
691 2208857, and 2225027 to TEXM.

692 **6 Authorship statement**

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

695 **7 Data accessibility statement**

696 No original data appear in this paper. Should the paper be accepted, all computer scripts  
697 supporting the results will be archived in a Zenodo package, with the DOI included at  
698 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).  
699

700 **Literature Cited**

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

- 730 Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying  
731 a new structured population model. *Ecology*, 81(3):694–708.
- 732 Elderd, B. D. and Miller, T. E. (2016). Quantifying demographic uncertainty: Bayesian  
733 methods for integral projection models. *Ecological Monographs*, 86(1):125–144.
- 734 Ellis, M. M. and Crone, E. E. (2013). The role of transient dynamics in stochastic popula-  
735 tion growth for nine perennial plants. *Ecology*, 94(8):1681–1686.
- 736 Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E., and Rees, M. (2022).  
737 A critical comparison of integral projection and matrix projection models for demo-  
738 graphic analysis: Comment. *Ecology*, 103(10):e3605.
- 739 Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modeling of Structured Popula-*  
740 *tions: A Practical Guide to the Integral Projection Model*. Springer, New York.
- 741 Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in  
742 animal populations. *Ecology*, 79:2531 – 2538.
- 743 Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New  
744 York, 2 edition.
- 745 Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed  
746 models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.
- 747 Héault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C. T., Wagner,  
748 F., and Baraloto, C. (2011). Functional traits shape ontogenetic growth trajectories of  
749 rain forest tree species. *Journal of ecology*, 99(6):1431–1440.
- 750 Jones, M. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96:761 – 780.
- 751 Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis.  
752 *The American Statistician*, 65(2):89 – 95.
- 753 Jones, O. R., Barks, P., Stott, I., James, T. D., Levin, S., Petry, W. K., Capdevila, P., Che-  
754 Castaldo, J., Jackson, J., Römer, G., et al. (2022). Rcompadre and rage—two r pack-  
755 ages to facilitate the use of the compadre and comadre databases and calculation of  
756 life-history traits from matrix population models. *Methods in Ecology and Evolution*,  
757 13(4):770–781.
- 758 Komsta, L. and Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and  
759 related tests. *R package version*, 14(1).

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

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

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

820 Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational  
821 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

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

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

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

860 
$$X_{\lambda, \tau} = \sinh(e^{-\tau} \operatorname{asinh}(Z) + \lambda). \quad (\text{S.3})$$

861 Thus, the two parameterizations are related by

862 
$$\delta = e^\tau, \epsilon = \delta\lambda = e^\tau\lambda. \quad (\text{S.4})$$

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

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

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

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

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

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

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

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

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

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

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

```

916 LogLik=function(pars,new.size,U){
917   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
918   mu = U%*%pars1;
919   sigma = exp(pars2[1]+pars2[2]*mu);
920   dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)
921 }
```

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

```

924 start=c(runif(ncol(U)), rep(0,3))
925 out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,
926   method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);
927 coefs = out$estimate; # parameters
928 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

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

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

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

$$953 \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

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

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

$$961 \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

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

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

$$967 \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

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

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

$$975 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.10})$$

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

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

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

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

```
996 # Variance-covariance matrices for intercepts and slopes
997 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
998 # Extract year-specific intercepts, center them to zero
999 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
1000
1001 # Estimate sigma^2
1002 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
1003           (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
1004
1005 # Shrink deviations from the mean
```

```

1006 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
1007
1008 # Do it all again for the slopes
1009 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
1010 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
1011           (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
1012 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

1013 The figure below shows the results for one artificial PSSP “data” set, having  $T = 22$   
1014 years and growth measurements on about 175 individuals/year on average. The true  
1015 random year effects (the ones used to generate the data) are recovered with good accu-  
1016 racy and no bias. In particular there is no sign of extreme values being pulled in too  
1017 far towards the mean, which would cause an S-shaped graph of estimated versus true  
1018 values.

