

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

Tom E.X. Miller^{*a} and Stephen P. Ellner^b

^aDepartment of BioSciences, Rice University, Houston, TX

^bDepartment 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

¹ **Abstract**

- ² 1. Integral Projection Models (IPMs) are widely used for studying the dynamics of
³ continuously size-structure populations. IPMs require a growth sub-model that
⁴ describes the probability of future size conditional on current size. Over the past
⁵ two decades, most IPM studies have assumed that this probability is normally-
⁶ distributed, despite repeated calls for non-Gaussian approaches that accommodate
⁷ skewness and kurtosis known to occur in size transition data.
- ⁸ 2. We provide a general workflow for modeling size transitions that accommodates
⁹ non-Gaussian growth patterns while retaining the desirable features (ecologically
¹⁰ important covariates and random effects) that Gaussian approaches typically pro-
¹¹ vide. Our approach emphasizes visual diagnostics of residuals from pilot Gaussian
¹² models and quantile-based metrics of skewness and kurtosis that vet the fit of the
¹³ Gaussian distribution and guide the selection of an alternative, if necessary. We
¹⁴ illustrate our methods by reanalyzing size transition data from our published IPM
¹⁵ studies, targeting a diversity of demographic quantities including population growth
¹⁶ rate, invasion wave velocity, and evolutionarily stable life history strategies.
- ¹⁷ 3. Across one coral and three plant case studies, skewness and excess kurtosis were
¹⁸ common features of size transition data and non-Gaussian growth models consis-
¹⁹ tently generated simulated data that were more consistent with the real data than
²⁰ pilot Gaussian models. However, in these case studies, the effects of “improved”
²¹ growth modeling on IPM results were generally modest, and differed in direction or
²² magnitude between different outputs from the same model.
- ²³ 4. Using tools that were not available when IPMs were first developed, it is now possi-
²⁴ ble to fit non-Gaussian models to size transition data without sacrificing ecological
²⁵ complexity; our worked examples demonstrate how, including open-access data and
²⁶ computing scripts. Doing so, as guided by careful interrogation of the data, will re-
²⁷ sult in a model that better represents the population for which it is intended.

²⁸ **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

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

202 where $m_k = \mathbb{E}(X - \bar{X})^k$ is the k^{th} central moment of a random quantity X and σ^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, it only takes a few outliers to produce
 208 a very inaccurate estimate. Figure 2 shows a simulated example, where the underlying

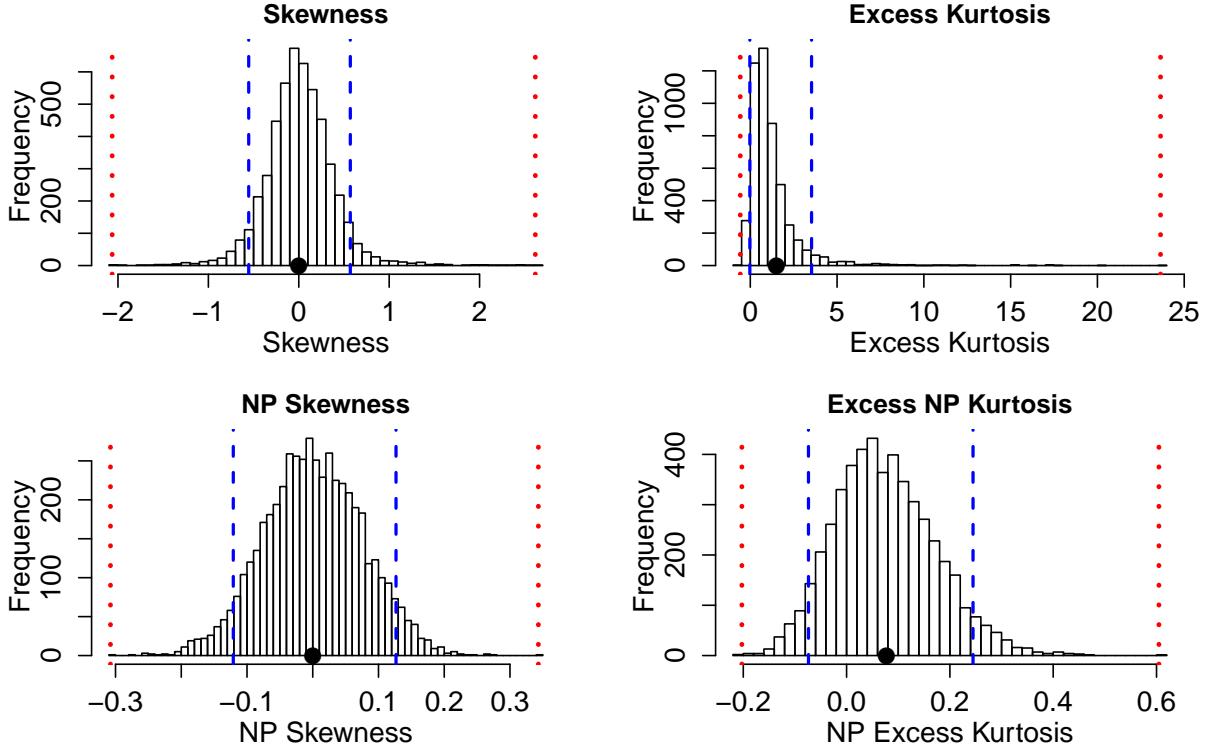


Figure 2: Histograms of skewness and kurtosis estimates using moment-based definitions, compared with the nonparametric measures. Histograms are based on 5000 replicate draws of a sample of 200 independent values from a t distribution with 8 degrees of freedom. Dotted red vertical lines mark the minimum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

“data” are a sample 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 good size 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_α denote the α 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 ± 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())
```

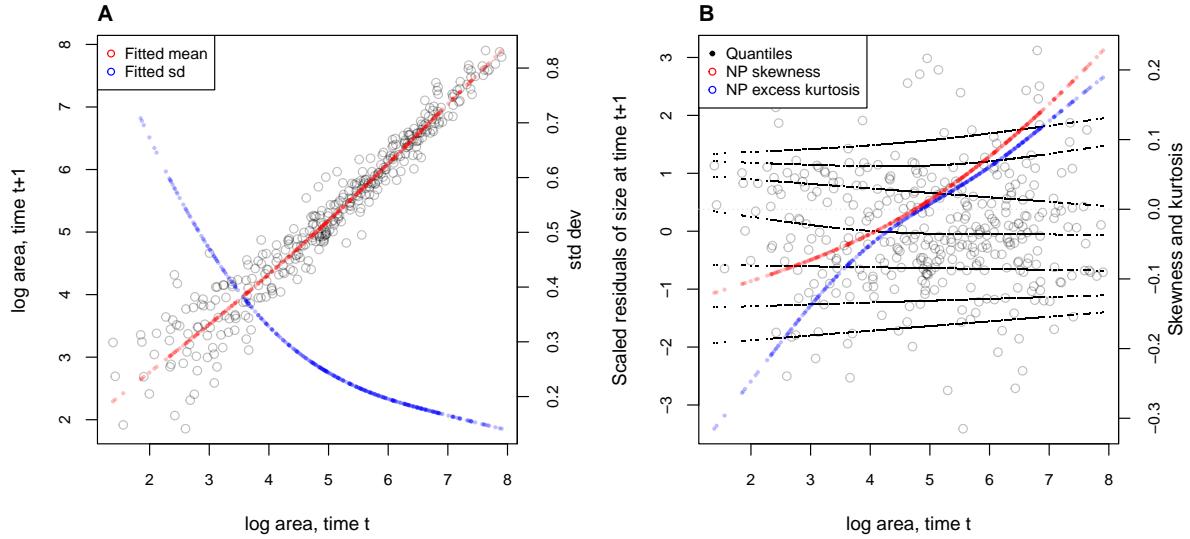


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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 blue curve) is visually nearly linear, but the fitted nonlinear spline is strongly favored over a linear model for the mean ($\Delta AIC \approx 9$). The spline for standard deviation σ versus initial size shows that smaller individuals exhibit greater variability in future size.

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

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

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

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

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

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

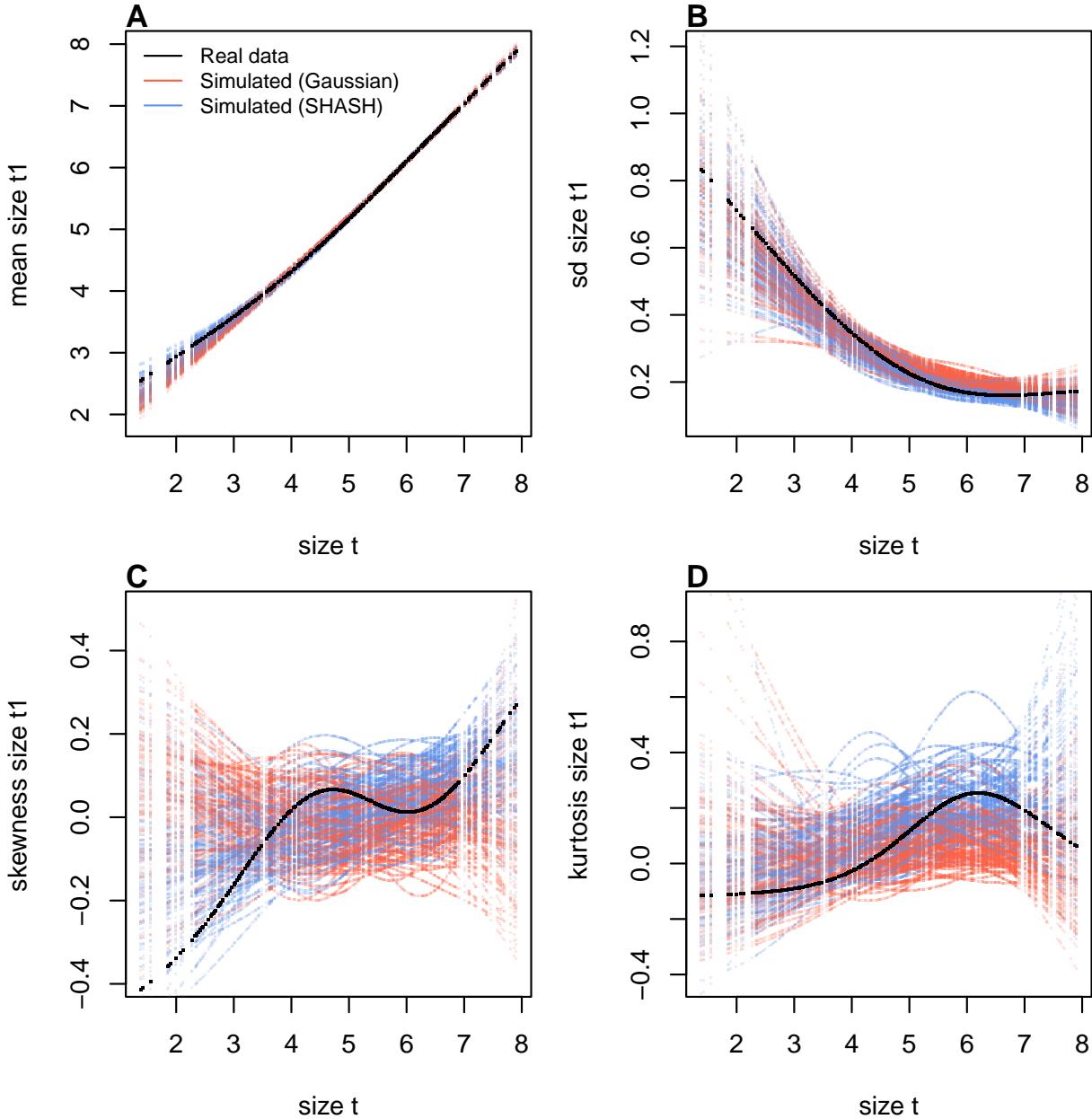


Figure 4: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `AkumalCorals_qgam.R`.

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

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

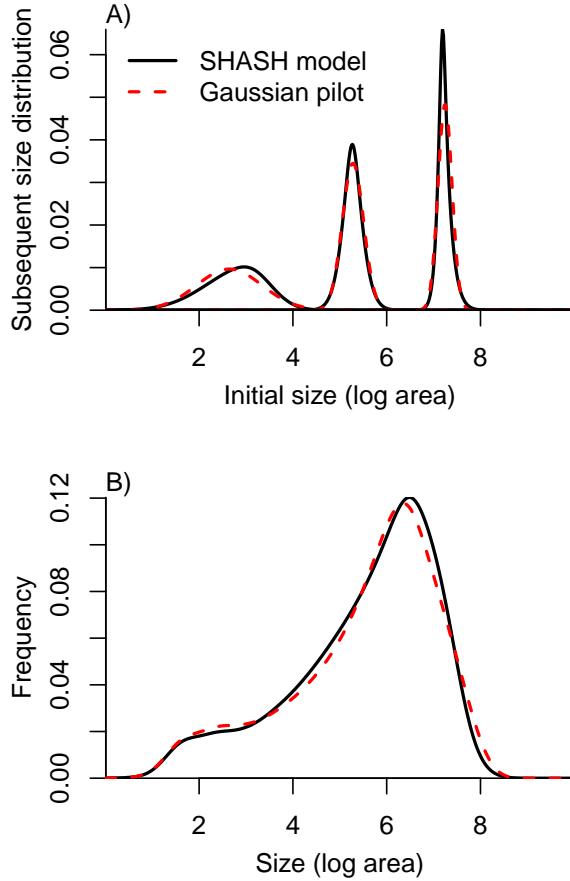


Figure 5: Comparisons between the fitted SEP1 growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year $t + 1$ for three different values of size in year t . The leftmost pair of curves are for initial size equal to the median size of a new recruit (data from (Bruno et al., 2011)); central pair is for the median initial size of uninfected individuals; rightmost pair are for the 95th percentile initial size for uninfected individuals. B) Steady-state size distributions resulting from a constant unit input of new recruits. As in Bruno et al. (2011) we assume that larvae are very widely dispersed, so that the number of recruits arriving at any one location is independent of the local population abundance or structure. Size distribution of recruits was described by a kernel density estimate based on the (sadly, only $n = 9$) measured sizes of known new recruits. Figure made by script AkumalCoralsIPMs.R.

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

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

352 2 Case study: tree cholla cactus, *Cylindriopuntia imbricata*

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

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

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

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

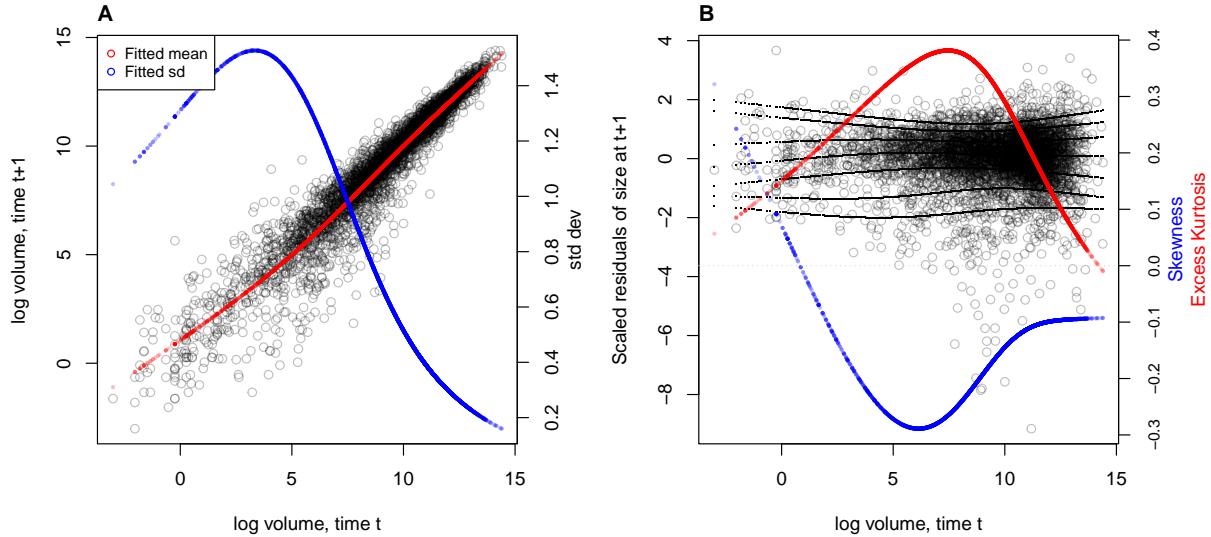


Figure 6: **A**, Size transition data for tree cholla cacti, *Cylindropuntia imbricata*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess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

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

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

  403 s(plot,bs="re") + s(year_t,bs="re")), # location
```

```

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

```

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

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

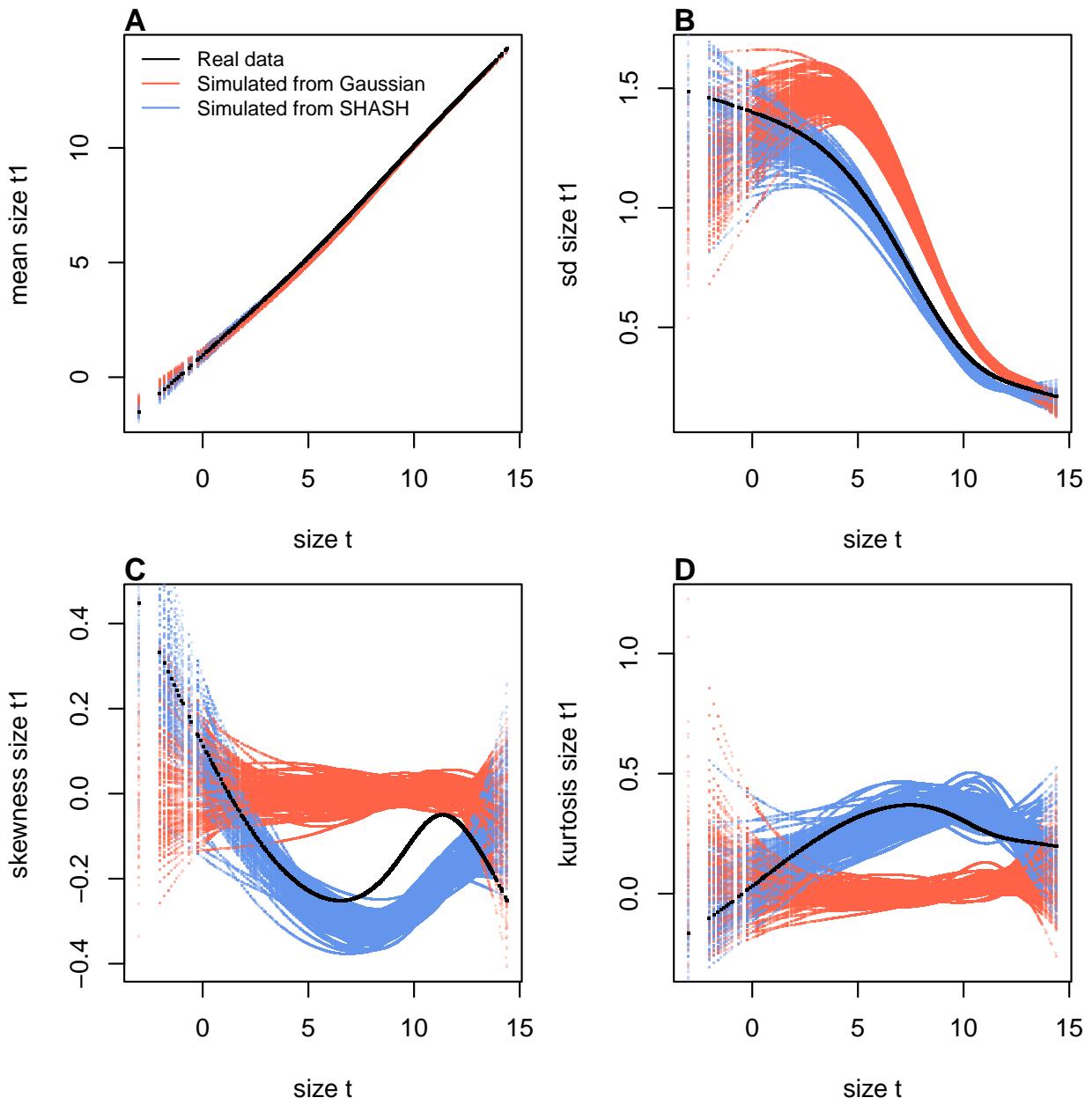


Figure 7: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

⁴³⁹ of one's analysis, modeling non-Gaussian size transitions with a Gaussian growth model
⁴⁴⁰ could bias results in either direction, or make no difference at all.

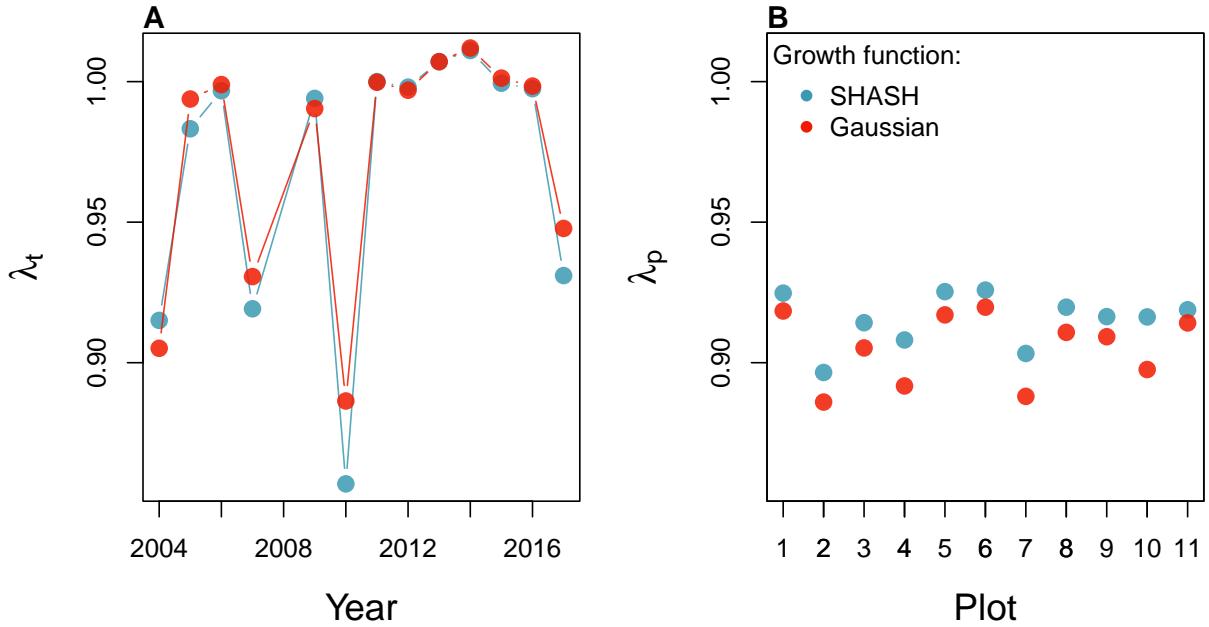


Figure 8: Temporal (A) and spatial (B) heterogeneity in fitness for the tree cholla cactus (*Cylindropuntia imbricata*) predicted by IPMs using Gaussian or SHASH growth models. Figure made by script `cactus_growth_modeling_qgam.R`.

441 3 Case study: creosotebush, *Larrea tridentata*

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

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

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

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

480 Unlike our previous case studies, here we have multiple fixed effects that may influ-
481 ence the variance of future size. In cases such as this, we recommend modeling variance
482 as a function of expected future size rather than initial size, as we did with the corals
483 and cacti. The expected (or “fitted”) values reflect the combined influence of all fixed
484 and random effects, and therefore implicitly account for multiple sources of variation in
485 the variance. While there are several convenient software packages for simultaneously
486 modeling Gaussian mean and variance as functions of independent variables (**mgcv** for
487 additive models as we saw above, **nlme** for linear models), **modeling variance as a func-**
488 tion of the mean is trickier because they cannot easily be fit simultaneously². Here we
489 us an iterative re-weighting approach – which is not elegant, but it works. For Gaus-
490 sian models, weights w_i can be used to indicate that the observations y_i vary in their
491 dispersion around the mean. In general, the iterative steps are:

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

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

$$y_i = \mu_i + \epsilon_i$$

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

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

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

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

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

$$y_i = \mu_i + \epsilon_i$$

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

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

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

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

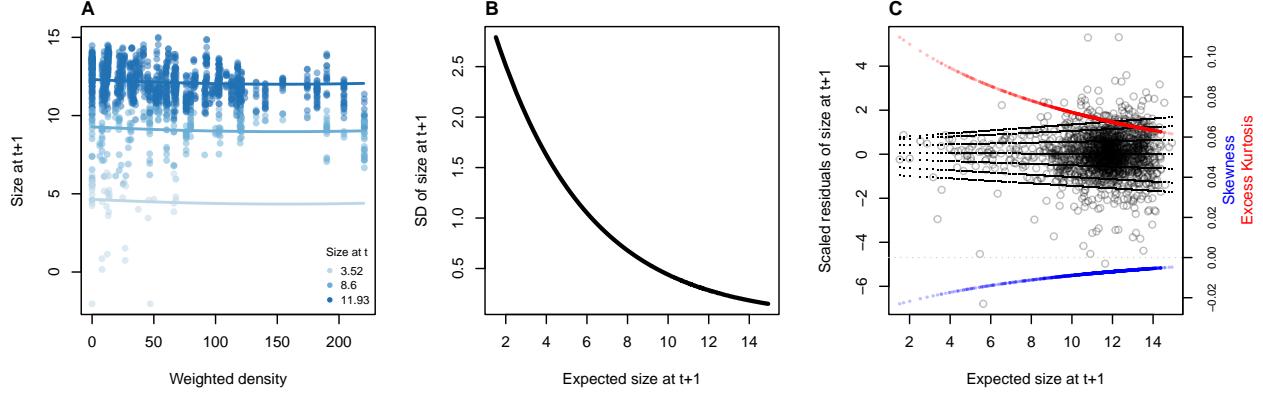


Figure 9: **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume (cm^3). **B**, Standard deviation of size at time $t + 1$ as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

parameter μ is the mean and scale parameter σ is the standard deviation (Rigby et al., 2019).

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

```

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

```

527 }

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

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

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

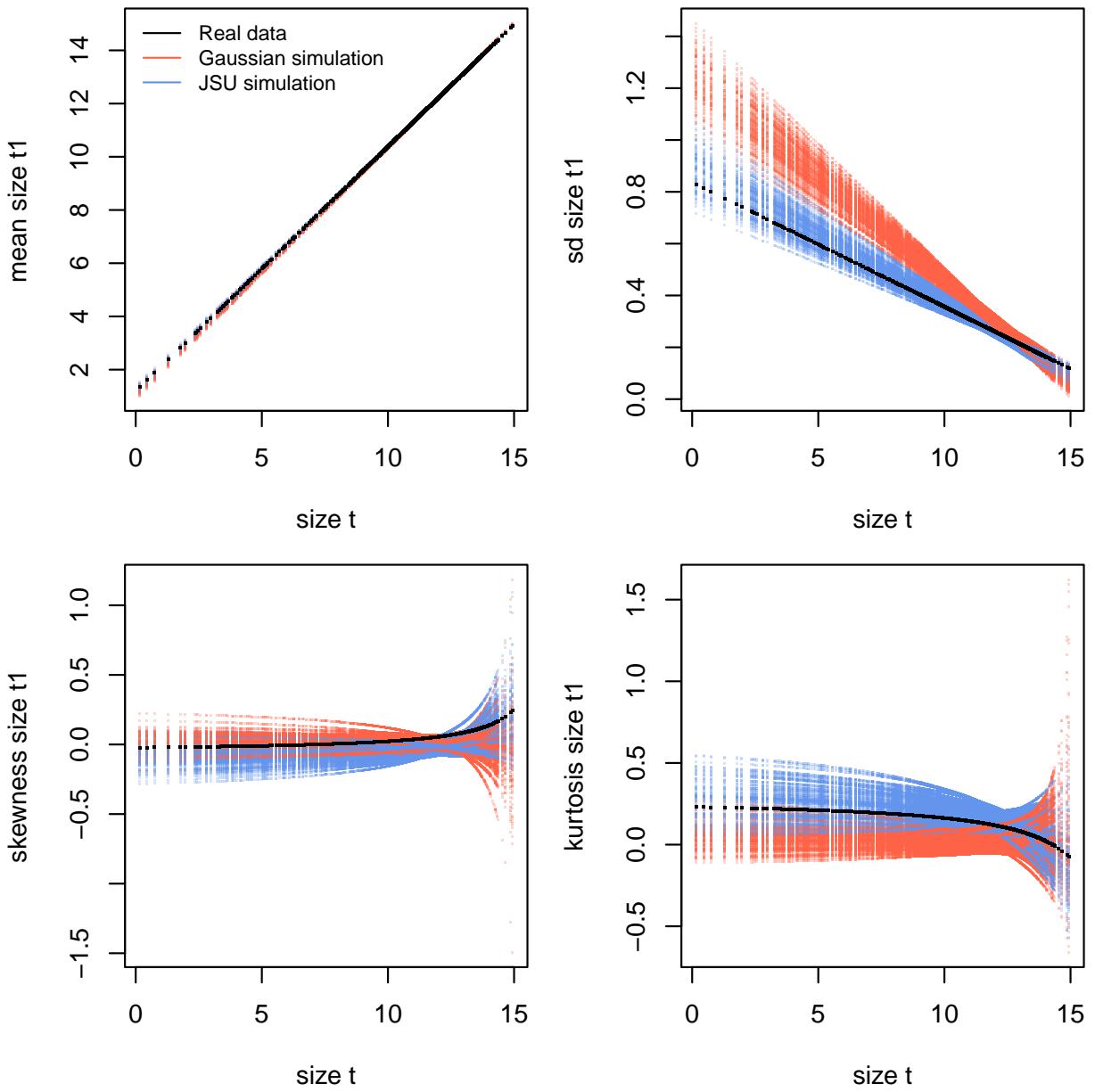


Figure 10: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `creosote_growth_modeling_qgam.R`.

558 4 Case study: lady orchid, *Orchis purpurea*

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

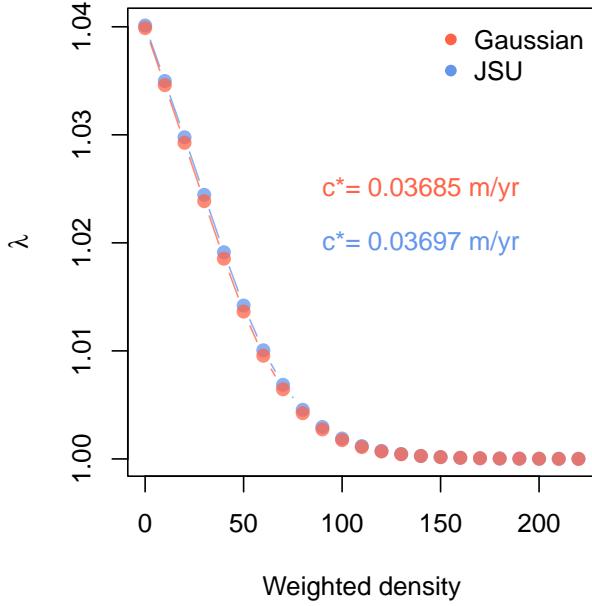


Figure 11: Density dependence in fitness (λ) and asymptotic velocity of the creosote encroachment wave (c^*) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ($\log(cm^3)$) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

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

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

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

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

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

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

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

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

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

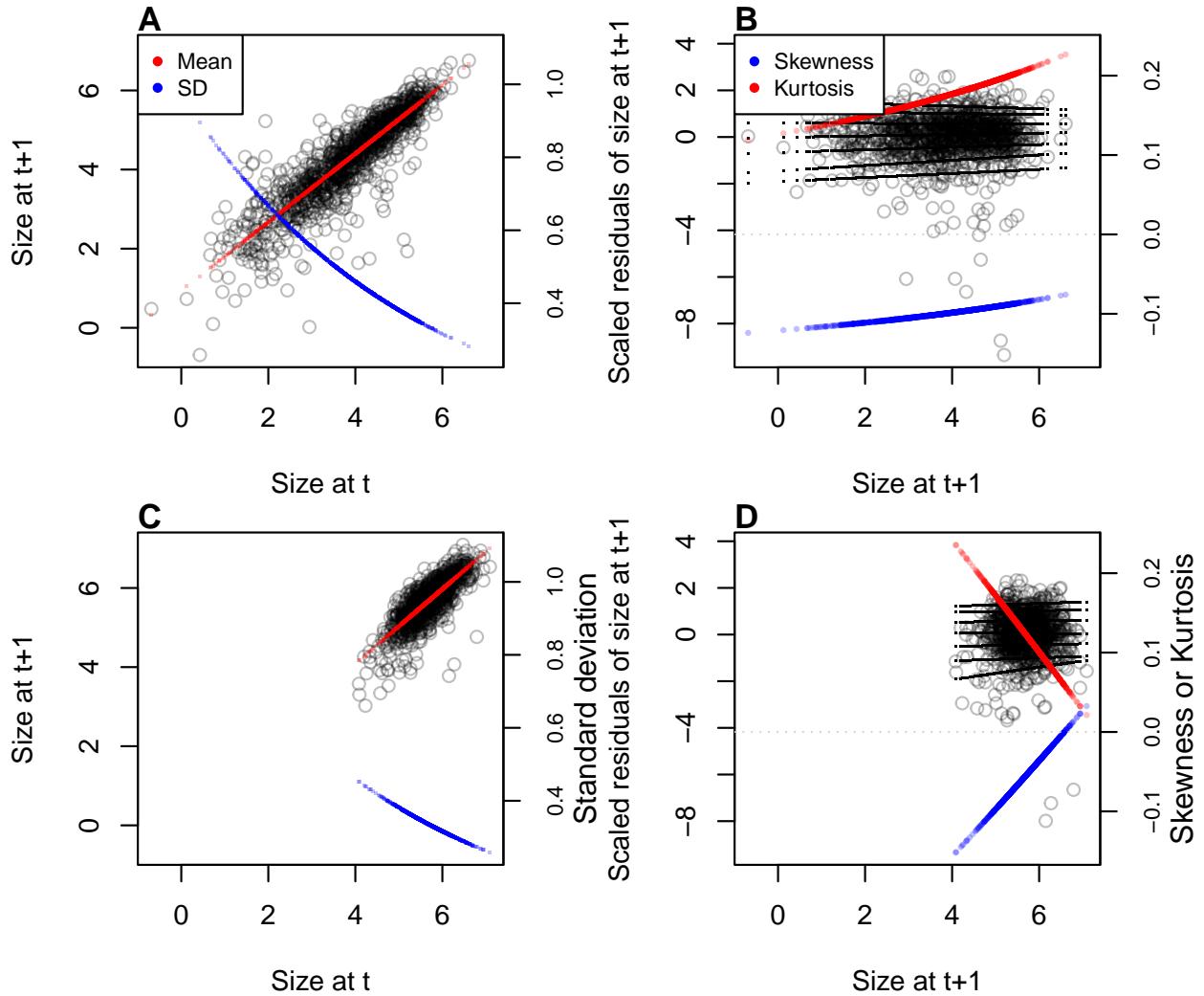


Figure 12:

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

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

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

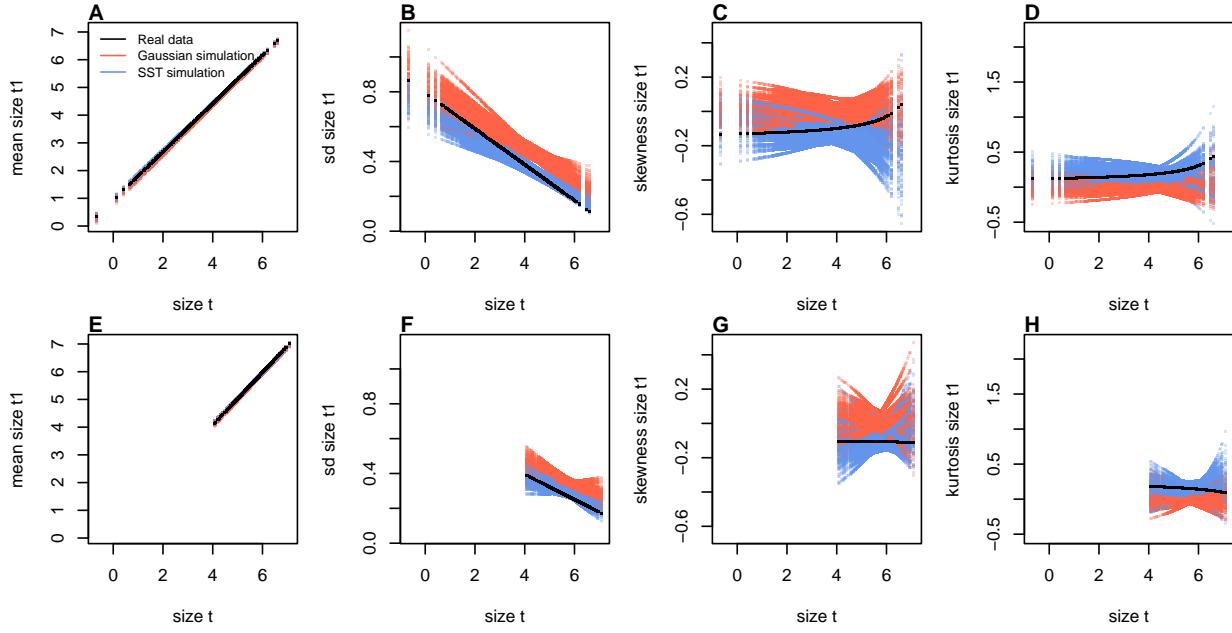


Figure 13: Comparisons between real orchid data and data simulated from Gaussian and skewed t growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (A-D) shows plants that were vegetative at the start of the transition year and bottom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

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

5 Discussion

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

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

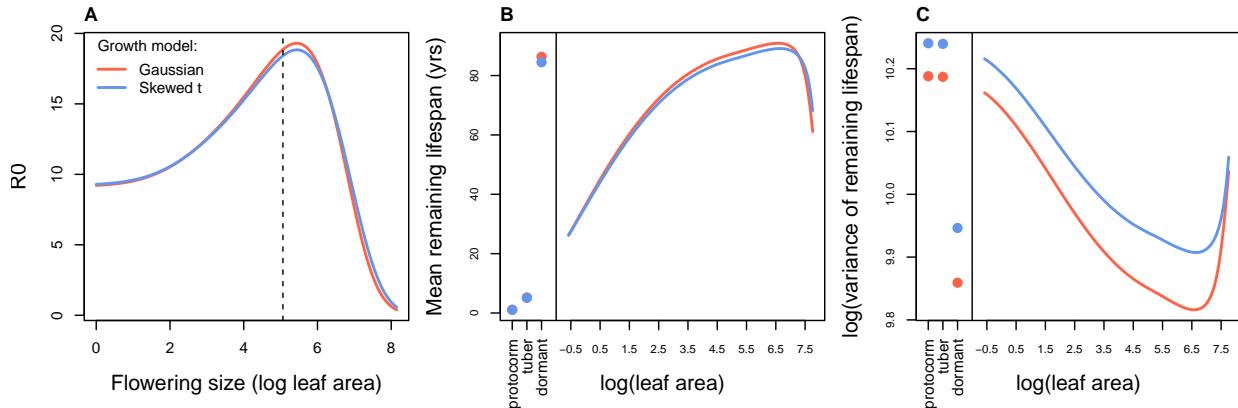


Figure 14: Orchid life history results from IPMs using Gaussian or skewed t growth models. **A**, Lifetime reproductive success (R_0) as a function of mean size of flowering. Dashed vertical line shows the observed mean flowering size. **B-C**, Mean and variance of remaining lifespan as a function of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

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

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

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

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

678 Some issues to be discussed.

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

683 Acknowledgements

684 This research was supported by US NSF grants DEB-1933497 to SPE and DEB-1754468,
685 2208857, and 2225027 to TEXM.

686 **6 Authorship statement**

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

689 **7 Data accessibility statement**

690 No original data appear in this paper. Should the paper be accepted, all computer scripts
691 supporting the results will be archived in a Zenodo package, with the DOI included at
692 the end of the article. During peer review, our data and code are available at [693 https://github.com/texmiller/IPM_size_transitions](https://github.com/texmiller/IPM_size_transitions).

694 **Literature Cited**

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

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

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

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

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

814 Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational
815 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 (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 ϵ , 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 μ and scale parameter σ 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 μ is not the mean, σ is not the standard deviation, ϵ is not the skew and δ is not the kurtosis. All else being equal, larger μ gives a larger mean, larger σ gives a higher standard deviation, higher ϵ gives higher asymmetry, and higher δ 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

844 to define a centered and scaled version in which μ and σ are the mean and standard
845 deviation.

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

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

854
$$X_{\lambda, \tau} = \sinh(e^{-\tau} \operatorname{asinh}(Z) + \lambda). \quad (\text{S.3})$$

855 Thus, the two parameterizations are related by

856
$$\delta = e^\tau, \epsilon = \delta\lambda = e^\tau\lambda. \quad (\text{S.4})$$

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

864 S.2 Estimating mixed-effects models using shrinkage

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

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

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

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

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

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

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

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

905 If there are T years, the matrix `U` specified in this way has $2T$ columns corresponding to
906 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:

```

910 LogLik=function(pars,new.size,U){
911   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
912   mu = U%*%pars1;
913   sigma = exp(pars2[1]+pars2[2]*mu);
914   dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)
915 }
```

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

```

918 start=c(runif(ncol(U)), rep(0,3))
919 out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,
920   method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);
921 coefs = out$estimate; # parameters
922 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 σ^2 . We also assume that the estimates \hat{a}_t are unbiased, that is

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

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

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

$$947 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

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

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

$$955 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

956 $\hat{\sigma}^2$ estimates the variance of the distribution from which the a_t are assumed to be drawn.

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

$$961 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

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

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

$$969 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.10})$$

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

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

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

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

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

```

1000 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
1001
1002 # Do it all again for the slopes
1003 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
1004 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
1005   (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
1006 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

1007 The figure below shows the results for one artificial PSSP “data” set, having $T = 22$
1008 years and growth measurements on about 175 individuals/year on average. The true
1009 random year effects (the ones used to generate the data) are recovered with good accu-
1010 racy and no bias. In particular there is no sign of extreme values being pulled in too
1011 far towards the mean, which would cause an S-shaped graph of estimated versus true
1012 values.

