

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.

³ 2.

⁴ 3.

⁵ 4.

⁶ **Keywords**

7 Introduction

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

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

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

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

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

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

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

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

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

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

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

88 A general workflow for better growth modeling

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

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

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

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

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

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

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

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

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

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

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

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

174 How should skewness and kurtosis be measured?

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

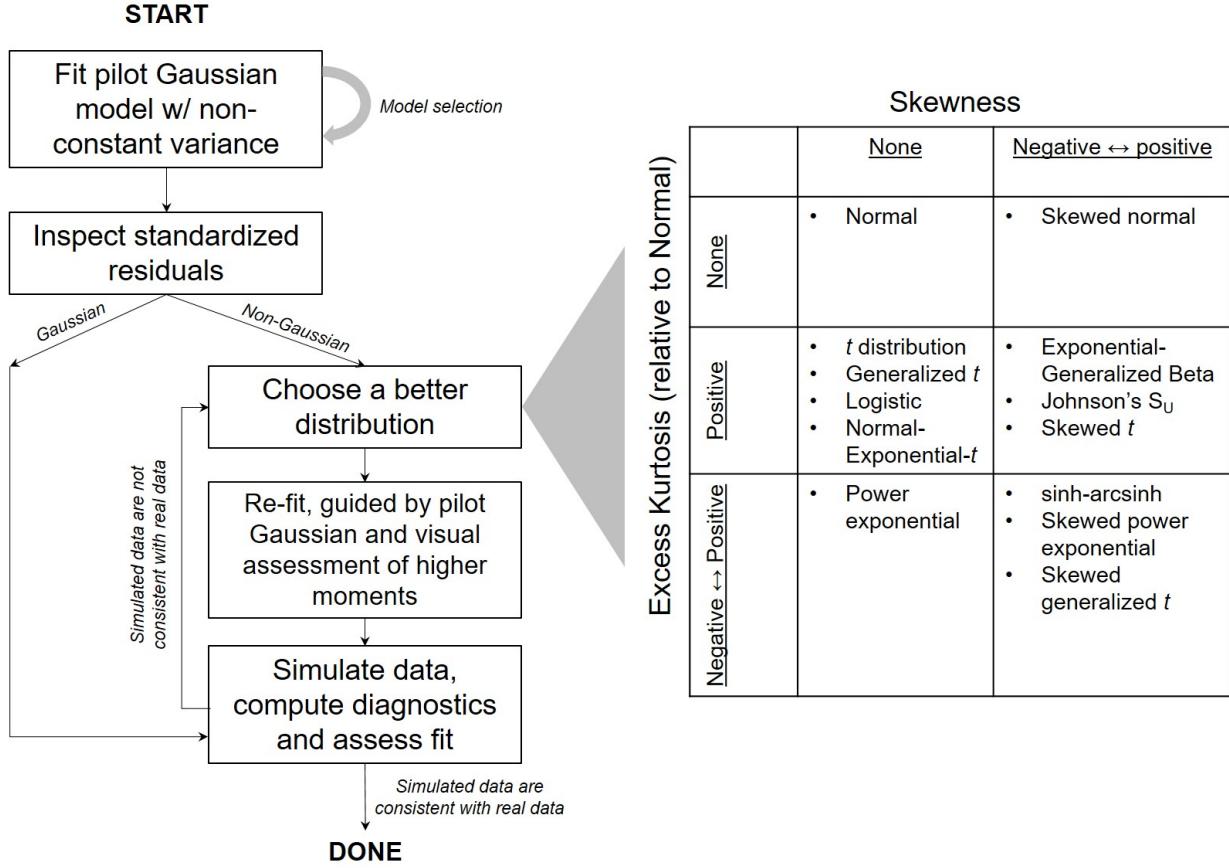


Figure 1: General workflow of recommendations for IPM growth modeling (left) and guide to common non-Gaussian distributions of size x for $x \in \mathbb{R}$ that can accommodate different combinations of skewness and kurtosis (right). All of these distributions (often including multiple versions or parameterizations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

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

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

¹⁸⁰ 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
¹⁸¹ variance (second central moment). A Gaussian distribution has zero skewness and zero
¹⁸² excess kurtosis.

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



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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

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

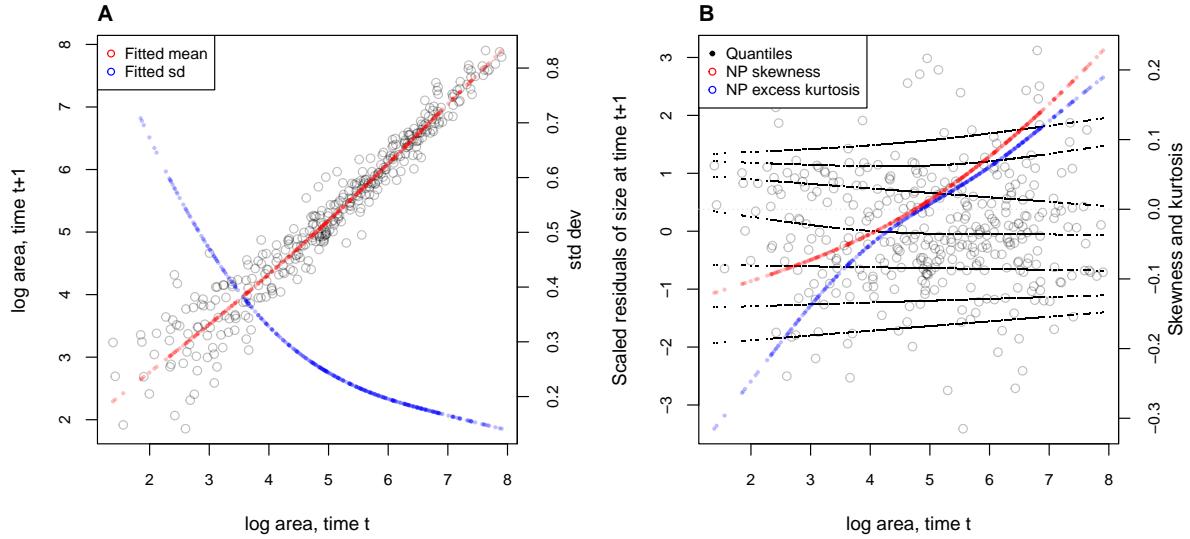


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis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

```

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

```

Fig. 3A shows the log-transformed data and Gaussian model. The mean function (solid blue curve) is visually nearly linear, but the fitted nonlinear spline is strongly favored over a linear model for the mean ($\Delta AIC \approx 9$). The spline for standard deviation 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 distribution, with smaller corals more likely to shrink than grow and larger corals more likely to grow than shrink. Kurtosis also changes direction over the size distribution, with smaller initial sizes having thinner tails and larger initial sizes having fatter tails than Gaussian. The fitted nonparametric moments suggest that the upper and lower tails of size transition proba-

278 bilities may differ by up to 20%, and the weight of the tails may be 20% greater or less
279 than Gaussian, depending on initial size – not overwhelming deficiencies, but not trivial
280 either. Are these deviations from normality severe enough to warrant a second, non-
281 Gaussian iteration of growth modeling? This question may be answered by simulating
282 data from the Gaussian model and examining whether key properties of the simulated
283 data are consistent with those of the real data – this is the ultimate litmus test for a
284 growth model’s adequacy and should be a standard element of IPM construction, in our
285 opinion. If the simulated data are not consistent with the real data, it is time to choose
286 a better distribution (Fig. 1). In this case, the negative skew at small sizes and excess
287 kurtosis observed at large sizes are more extreme than what occurs across 100 random
288 iterations of data simulation (Fig. 4), suggesting that, for at least some parts of the size
289 distribution, a non-Gaussian model would better capture size transitions.

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

```
294 fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0,k=4), # <- location  
295 ~ s(logarea.t0,k=4), # <- log-scale  
296 ~ s(logarea.t0,k=4), # <- skewness  
297 ~ s(logarea.t0,k=4)), # <- log-kurtosis  
298 data = XH, family = shash, optimizer = "efs")
```

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

305 What, then, have we gained by fitting a better growth model? Fig. 5A compares
306 the predicted distributions of subsequent size in the fitted model and Gaussian pilot
307 models, for the median size of a new recruit (leftmost pair of curves), the median initial
308 size (central curves), and the 95th percentile of initial size in the data (rightmost curves).
309 The differences are small, and most pronounced for the smallest size, where recruits
310 are predicted to grow slightly larger under the SHASH model than the Gaussian model.
311 The direction of this difference was surprising, since the SHASH accommodates negative
312 skew at small sizes in the data. However, in modeling skew appropriately, the SHASH

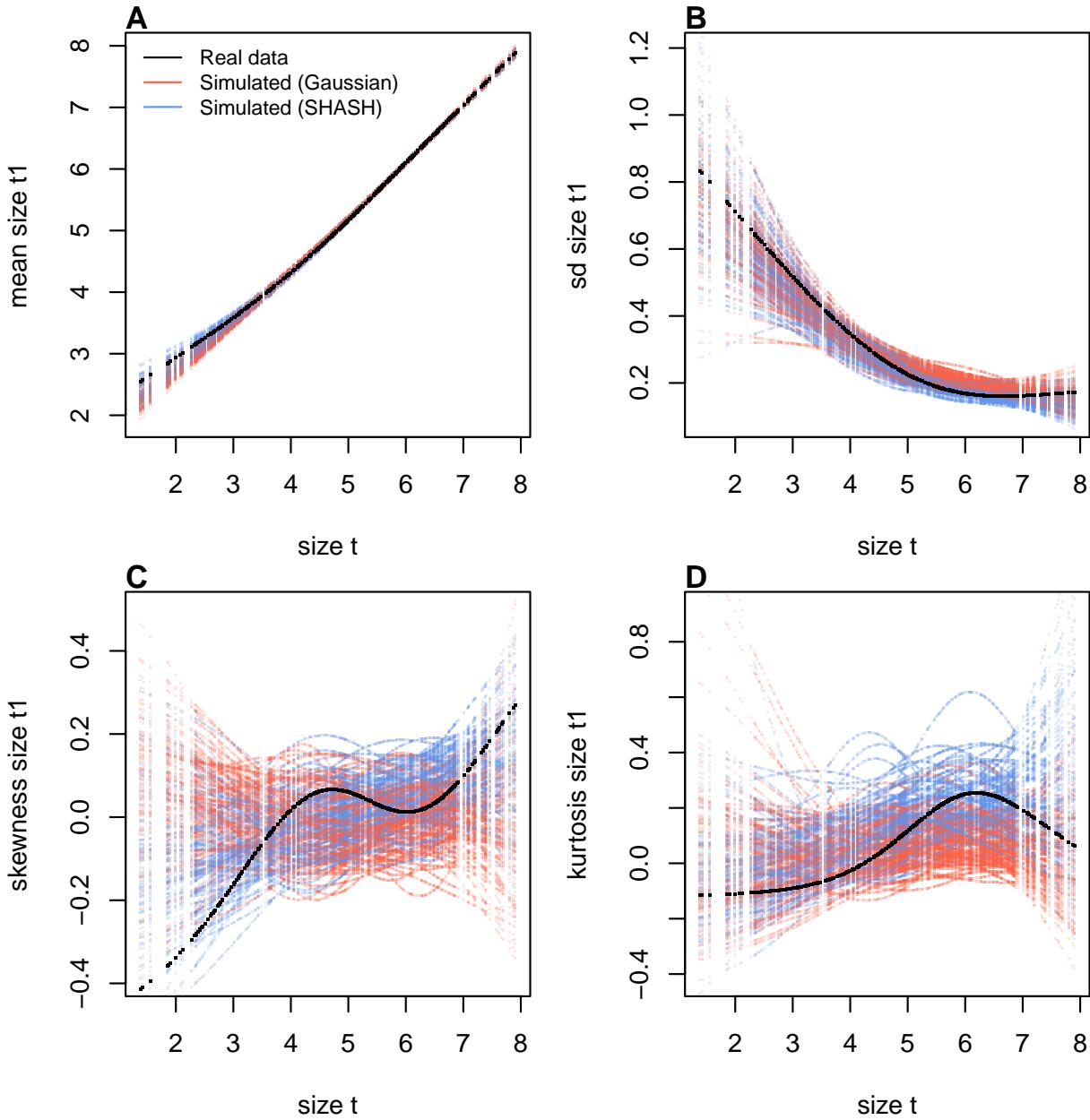


Figure 4: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Figure made by script `AkumalCorals_qgam.R`.

model also gives a better prediction for mean growth at small sizes than the Gaussian model, whose mean is biased downward by negative skew (Fig. 4A)³. 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

³...Contradicting the earlier assertion that parameter estimates from Gaussian models are robust to deviations from normality!

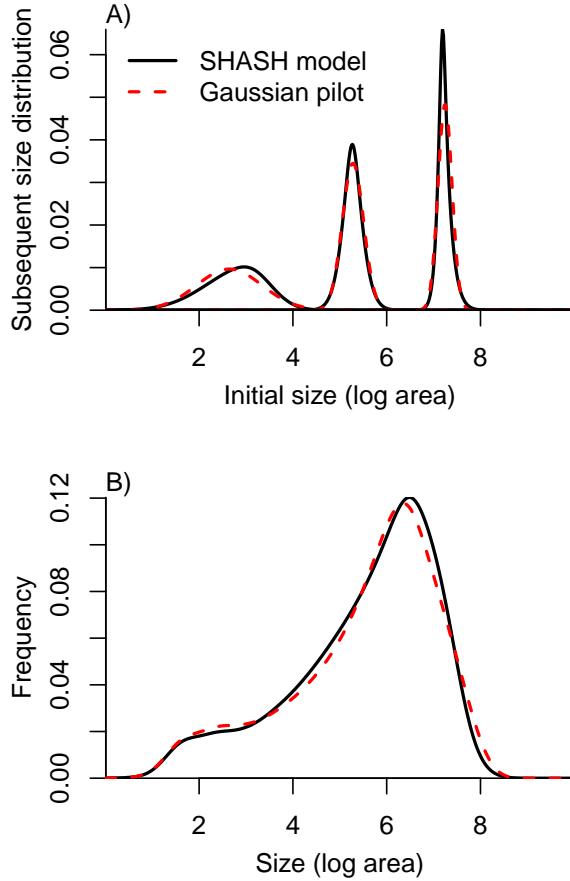


Figure 5: Comparisons between the fitted SEP1 growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year $t + 1$ for three different values of size in year t . The leftmost pair of curves are for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

317 distributions resulting from a constant unit input of recruits. Again, the differences are
 318 very subtle. Finally, the Gaussian and SHASH growth models predict very similar mean
 319 life span (17.7 and 17.9 years, respectively). From these outputs, there is little evidence
 320 that improved modeling of coral growth meaningfully improved biological inferences
 321 from the IPM; one could argue that it was not worth the trouble.

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

332 2 Case study: cactus, *Cylindriopuntia imbricata*

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

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

358 2.1 A pilot Gaussian model

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

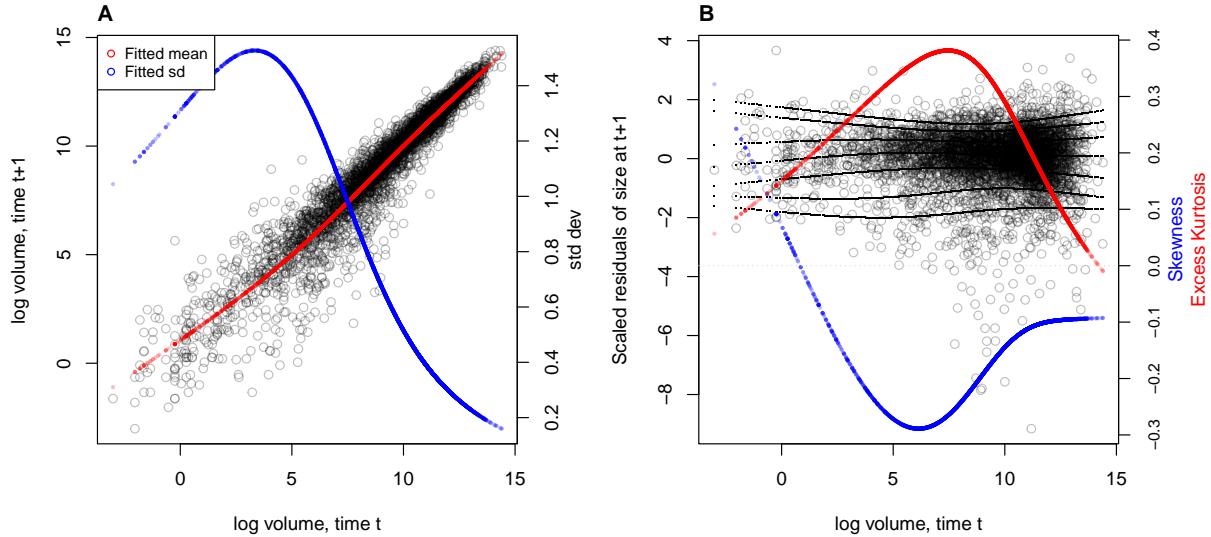


Figure 6: **A**, Size transition data for tree cholla cacti, *Cylindropuntia imbricata*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess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

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

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

376 2.2 An improved growth model

377 To better capture size transitions, we need a distribution with negative skew and positive
 378 excess kurtosis, but both of which may be negligible at some sizes. Appropriate candi-
 379 dates include Johnson's S_U distribution, which is limited to positive excess kurtosis, and
 380 the sinh-arcsinh (SHASH) and skewed power exponential distributions, which can vary

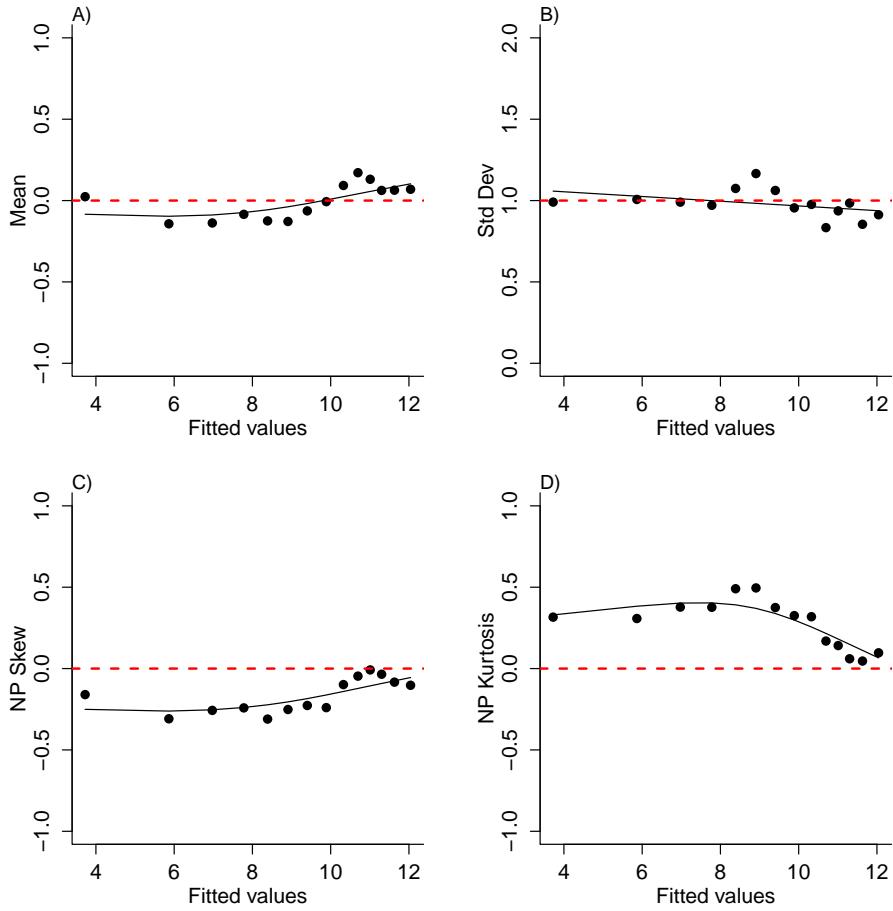


Figure 7: Rolling-moment analysis of standardized residual from a Gaussian fit to the cactus growth data. Red dashed lines indicate the Gaussian expectation. These data show skewness and kurtosis that deviate from Gaussian.

381 from leptokurtic to platykurtic (Fig. 1). As above, we divided the standardized residuals
 382 into discrete bins of initial size and fit competing distributions to each data subset using
 383 our function `gamlssMaxlik()`. We found that Johnson's S_U distribution (JSU in `gamlss`)
 384 was favored over most of the size distribution, particularly the larger sizes. However, at
 385 smaller sizes, the SHASH distribution was favored. They are both four-parameter dis-
 386 tributions but the SHASH is more flexible than the JSU, capable of capturing a greater
 387 range of possible kurtosis for a given amount of skew⁴. We ultimately settled on the
 388 SHASH distribution, but only after first trying the JSU and proceeding through the
 389 workflow (Fig. 1). We were unsatisfied with the correspondence between real and simu-
 390 lated data in the final step, so we followed the re-iteration loop back to "choose a better
 391 distribution", this time the SHASH. To keep this section concise, we do not fully nar-

⁴This comes from Steve's *NPSkewKurtosisRanges.pdf*

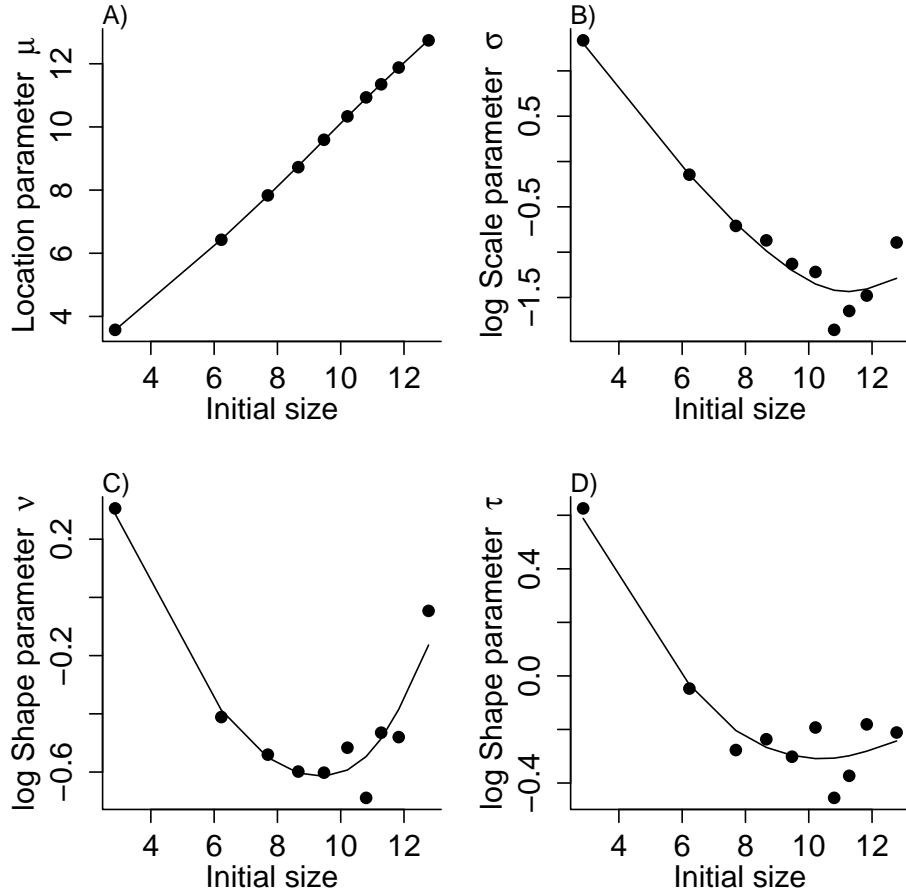


Figure 8: Parameter estimates of the SHASH distributions fit to cactus growth data in discrete bins of initial size.

rate that re-iteration loop, but we expect it will be a common feature of many growth analyses.

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

```

397 LogLik=function(pars,response,U){
398   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
399   mu = U%*%pars1;
400   val = dJSU(x = response,
401             mu=mu,
402             sigma = exp(pars2[1] + pars2[2]*x + pars2[3]*x^2),
403             nu = pars2[4] + pars2[5]*x + pars2[3]*x^2,
404             tau = exp(pars2[6] + pars2[7]*x + pars2[8]*x^2), log=T)

```

```
405     return(val)  
406 }
```

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

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

411 We can use the ‘shrinkage’ methods described in Appendix S.1 to estimate the vari-
412 ance terms associated with year and plot effects, even though they are fit here as fixed
413 effects. There is a tedious but not insurmountable complication that can arise in models
414 with multiple random effects, as in this one. By supressing the intercept in the model
415 matrix (the zero after the tilde), we get parameter estimates for each year rather than
416 offsets relative to a baseline year. However, because there is the additional effect of plot,
417 all of the year effects are conditioned on the first level of the plot factor variable (plot 1)
418 and the remaining plot coefficients are offsets of plot 1. The parameter estimate for plot
419 1 is therefore the mean of the year estimates, but there is no way to calculate its standard
420 error. We estimate the plot variance **excluding plot 1**⁵, a possible source of bias the vari-
421 ance estimate. The random effects estimated via the shrinkage method correspond well
422 with those from the original Gaussian mixed model, particularly for year effects (Fig.
423 9A). Plot effects were less tightly correlated but there was also less variance across plots
424 than there was across years (Fig. 9B).

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

432 2.3 Consequences of improved growth modeling for IPM predictions

433 Finally, we can ask how the improvements we have made to modeling cactus growth af-
434 fect inferences from the full IPM. Details of IPM construction and analysis are provided

⁵ *A very unsatisfying feature of this analysis. I've tried flipping the order so that plots are conditioned on year 1, but when I do it this way σ^2 for plot is negative. Not totally unreasonable because the plot effects are small relative to year effects, but still annoying. Steve has mentioned that we might get around this by abandoning the `model.matrix` function but I don't see how. We cannot get unconditional estimates of plot and year effects without adding a parameter to the model, and that parameter is unidentifiable.*

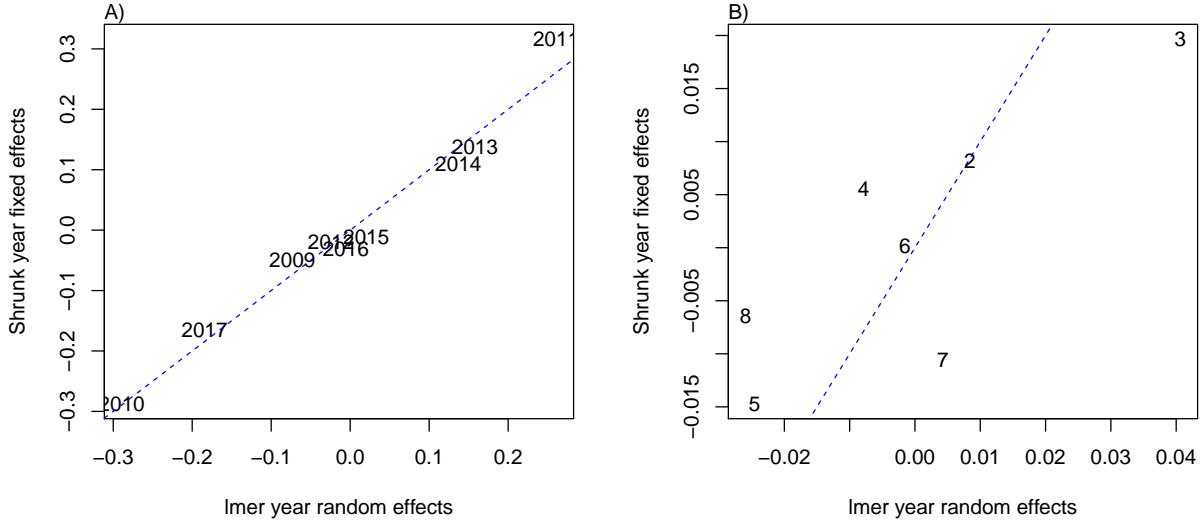


Figure 9: Comparison of year (A) and plot (B) random effects from ‘shrinking’ fixed-effect estimates (y-axes) vs. `lme4` mixed models (x-axes) for cactus data set.

in Appendix #⁶. Our previous IPM analyses of this system have consistently shown asymptotic population growth rates below replacement-level ($\lambda < 1$). Our new results with the improved SHASH growth kernel are qualitatively consistent with these results. However the SHASH growth kernel predicts a much faster rate of decline – a 9% difference in annual change – than the Gaussian, holding all else equal ($\lambda_{SHASH} = 0.901$ vs. $\lambda_{Gaussian} = 0.995$). The SHASH also predicts a very different and much smaller stable size distribution (SSD) than does the Gaussian (Fig. 11). The observed size distribution falls in between the two predictions, suggesting that, either way, this population is not at its stable distribution. The difference in predictions of the two IPMs is driven by contrasting size transitions of large plants (Fig. 12). This contrast is driven by skew and kurtosis, since the mean and variance of the SHASH and Gaussian models were nearly identical for large sizes (Fig. 10). The Gaussian growth kernel over-predicts size transitions at large sizes, and this leads to the larger median size at SSD and also the smaller peak corresponding to new recruits, since reproductive output is strongly size-dependent. This peak disappears from the SHASH SSD because reproductive output is strongly size dependent, and the SHASH growth kernel predicts that transitions to the largest sizes are less likely.

⁶Need to do. Maybe we can have one appendix for all the IPMs.

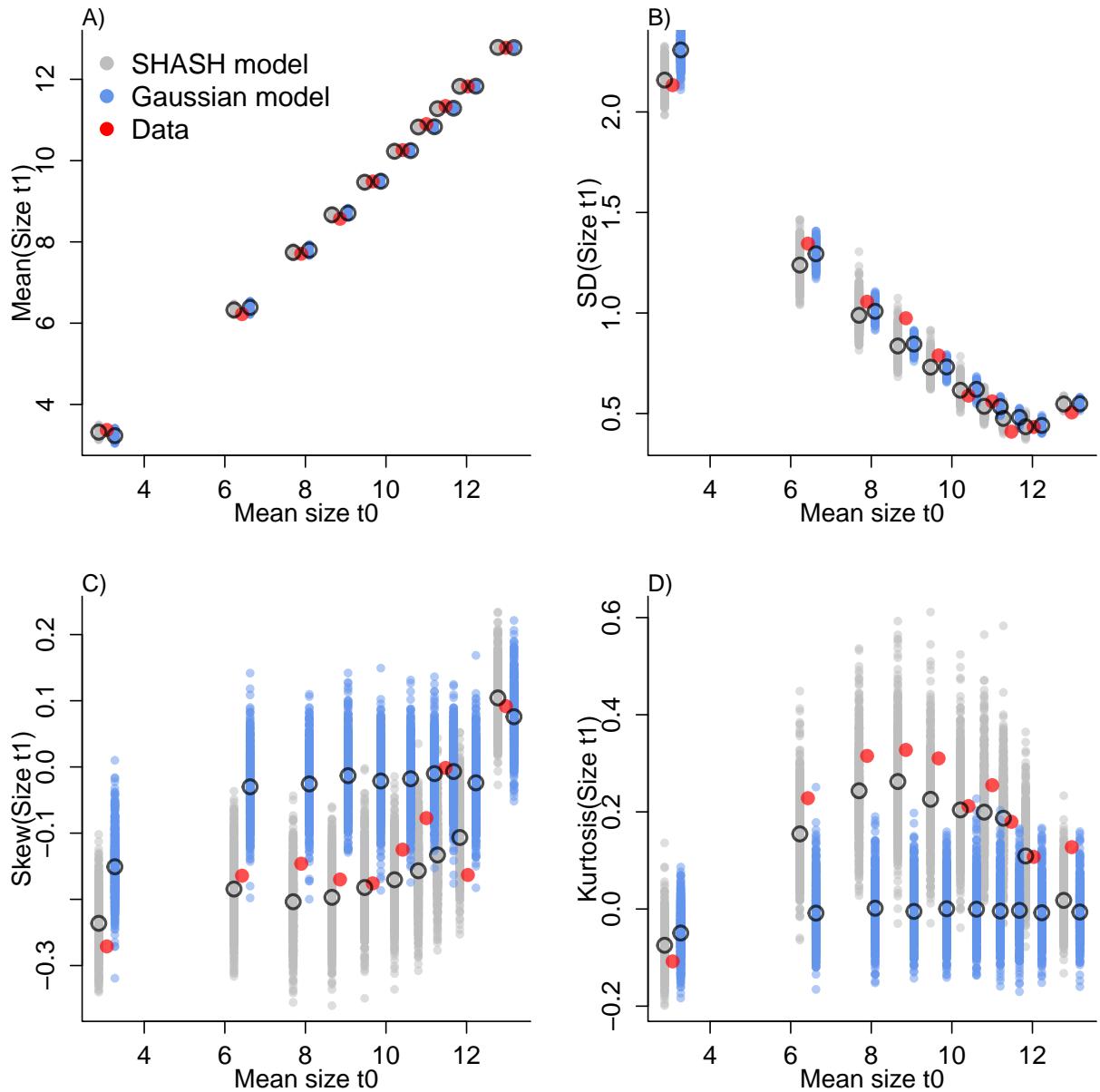


Figure 10: Simulated data from SHASH and Gaussian growth kernels compared to cactus growth data to which the models were fit.

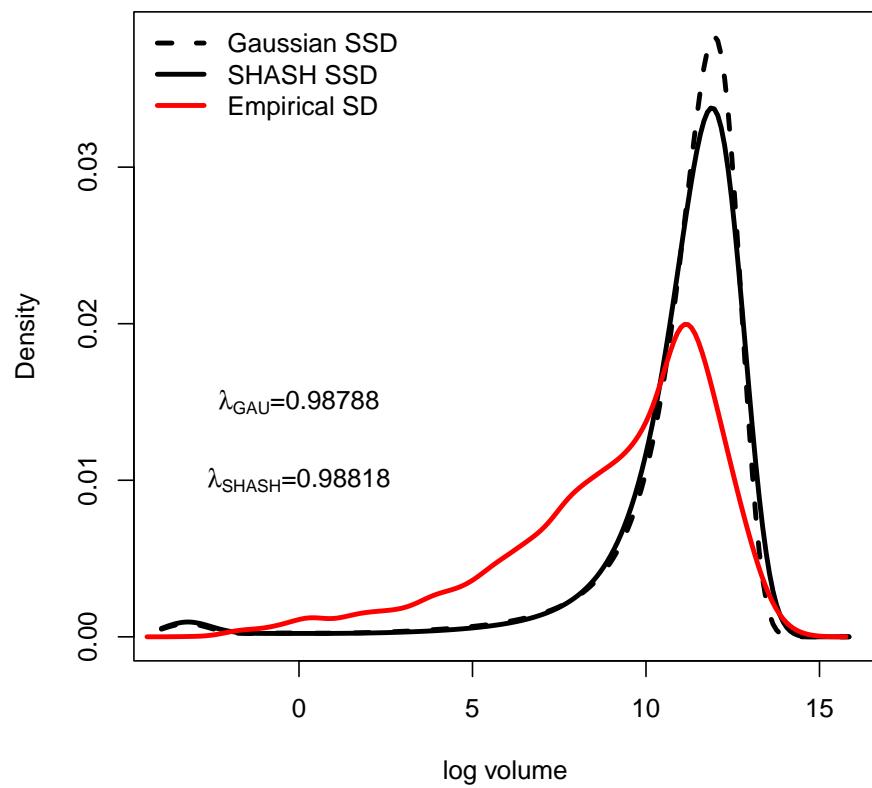


Figure 11: Predicted and observed size distributions of tree cholla cactus.

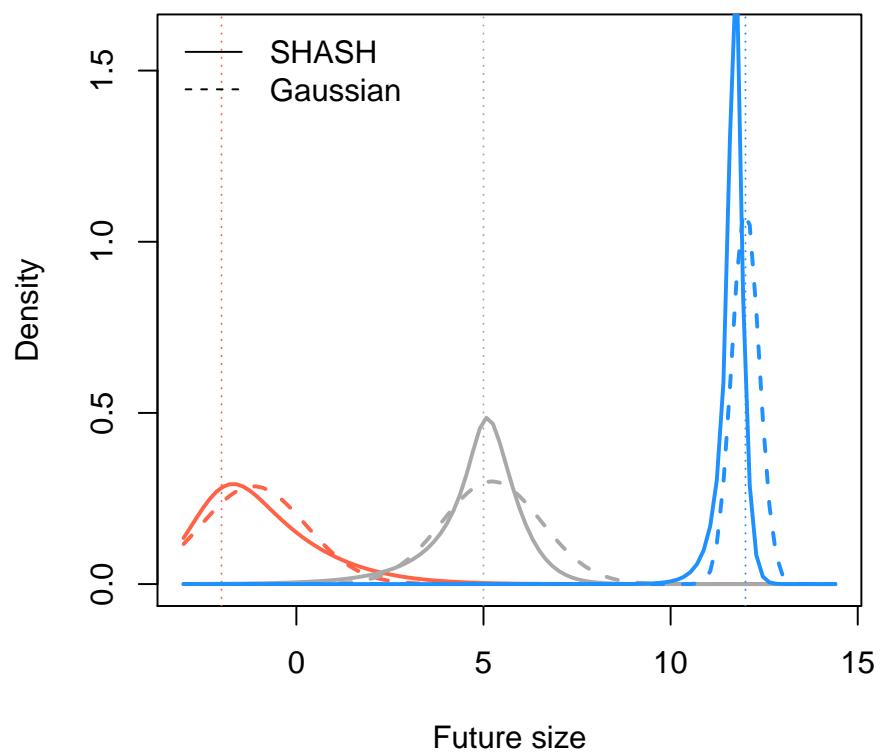


Figure 12: Predicted probability of future size given three initial sizes indicated by vertical lines.

452 **3 Case study: bunchgrass, *Pseudoroegneria spicata***

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

457 We used the most recently curated version of the data (Adler et al., 2018, at
458 doi.org/10.5061/dryad.96dn293), both legacy (22 annual transitions between 1926 and
459 1957) and modern (8 annual transitions from 2008 to 2016, excluding moisture manip-
460 ulation treatments). We excluded seedlings, which require separate models (Chu and
461 Adler, 2014, 2015; Snyder and Ellner, 2018), and individuals mapped as “too small to
462 measure” that should be modeled separately as a discrete size category (though in the
463 past we have not done that). The measure of plant size was log basal cover.

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

472 Results:

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

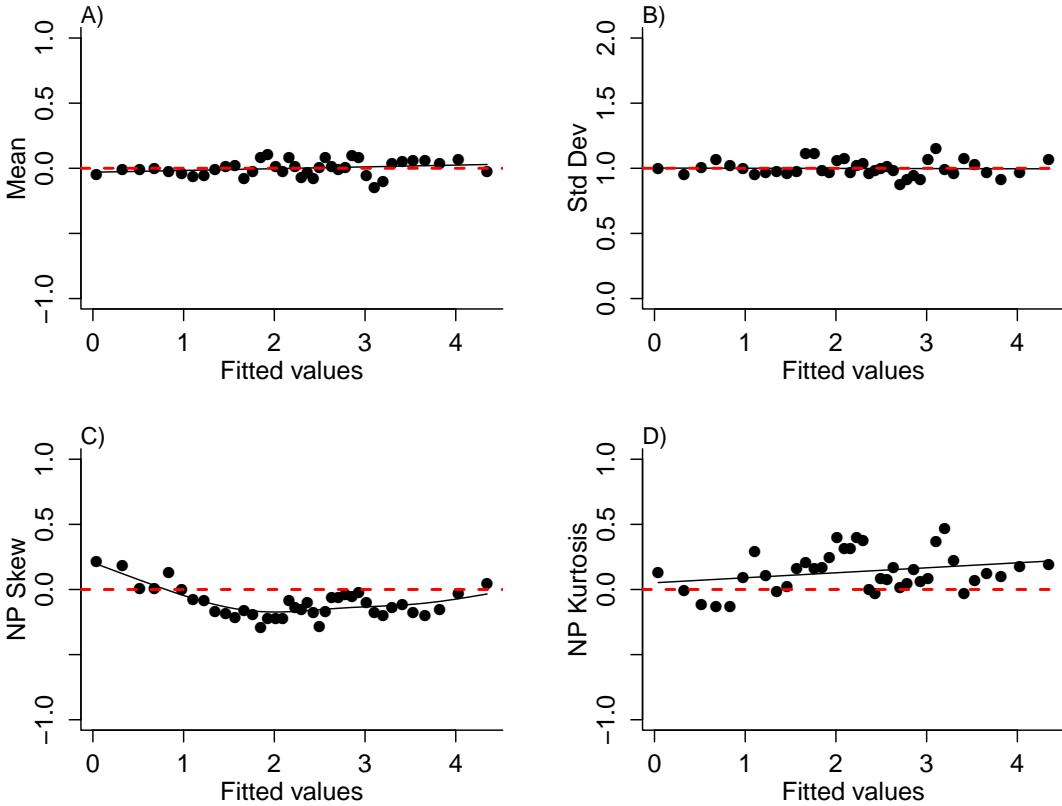


Figure 13: Rolling moments of scaled residuals from the pilot model, as a function of fitted values, for *P. spicata*. The dashed red lines are the value expected if the pilot model fits the data.

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

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

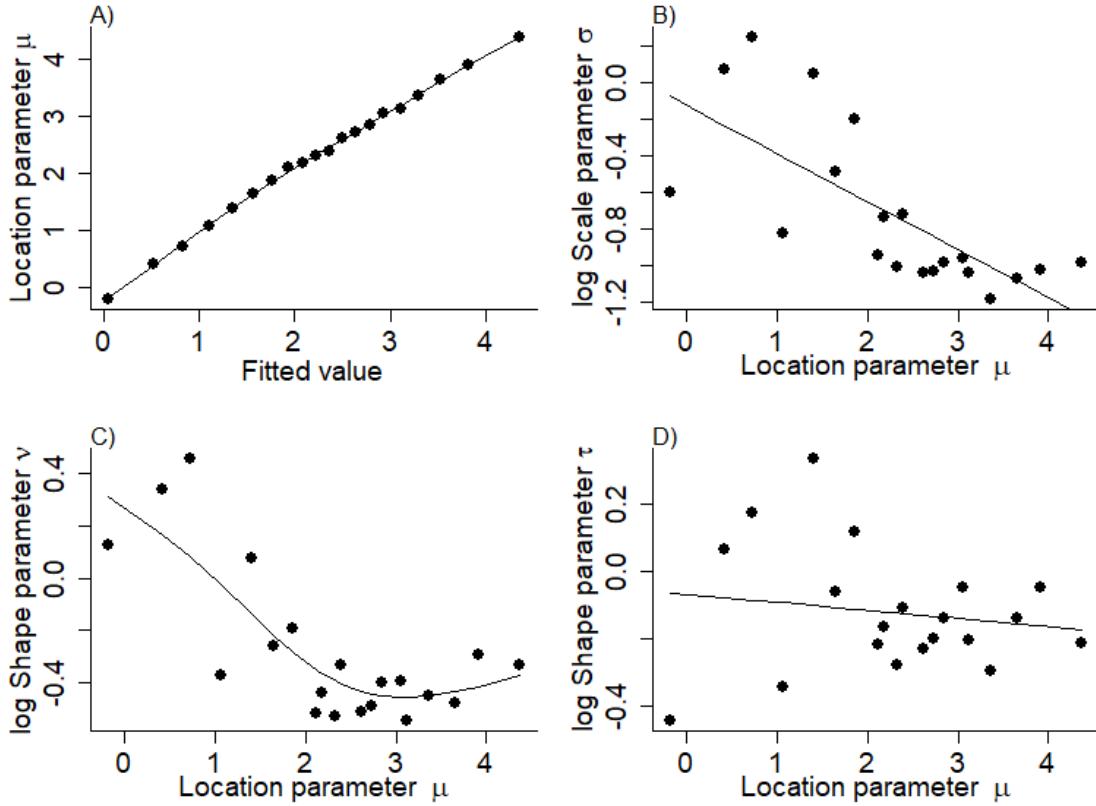


Figure 14: Binned data SHASH parameters plot for *P. spicata*.

- 500 5. What distribution families can accomodate the features in Fig. 13? We need at least
 501 4 parameters (to allow skew and excess kurtosis), and both postive and zero excess
 502 kurtosis must be possible. In JSU and skew t and that makes fitting problematic,
 503 because zero excess kurtosis only occurs as a limit, not at any actually possible
 504 parameter values ($df \rightarrow \infty$ for t , and $\tau \rightarrow 0$ for JSU). Comparison was therefore
 505 limited to SHASH, SHASHo, SEP1, SEP3, SEP4 – not SEP2 for the reasons noted above.
 506 The winner was SHASH.
- 507 6. Now we need to model the size-dependence of the SHASH distribution parameters.
 508 In the pilot model, σ as a function of fitted value was strongly favored or σ as a
 509 function of initial size. We retain that structure in the SHASH model. The data were
 510 therefore binned based on fitted values of the pilot model, and a SHASH distribution
 511 was fitted to each bin. fig. 14. Panel A), μ as a function of the fitted values, doesn't
 512 really inform the modeling, but it shows that it is safe to regard μ as equivalent to the
 513 "fitted value" in the pilot model. The other plots tell us how the other distribution

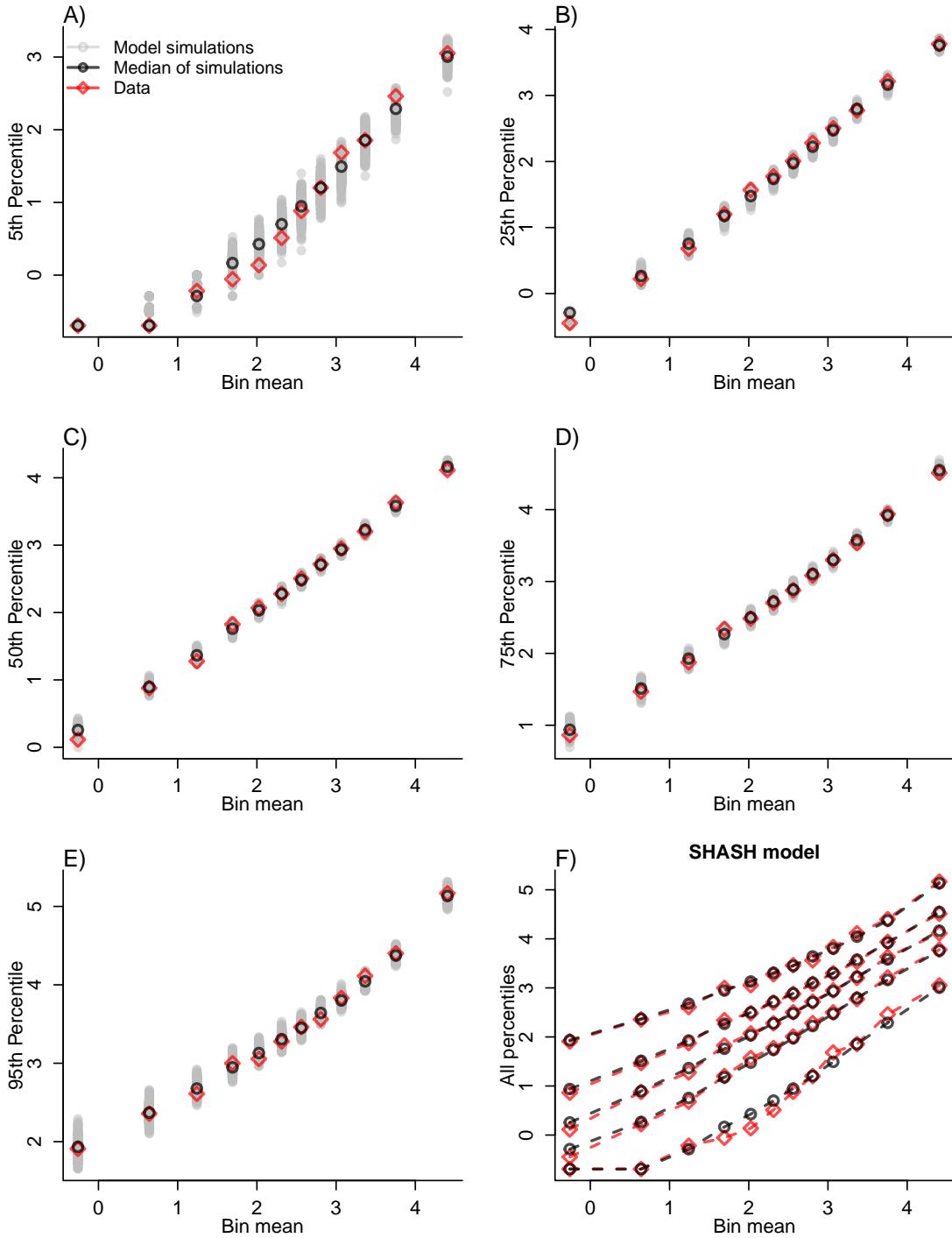


Figure 15: Binned data comparison of distribution quantiles between simulations of the fitted SHASH model (grey, black) and the actual data (red) for *P. spicata*. Individuals were binned based on their initial size.

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

514

515

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

- 516 7. The model is easy to fit by ML. Does it do a good job of fitting the data? The
517 comparison based on binned quantiles (fig. 15 looks pretty good except that the
518 variation in the 5th percentile of the data is more wiggly than the 5th percentile
519 of the model. To eliminate this imperfection, we could perhaps try higher-order
520 polynomials for ν or τ .
- 521 8. More importantly, the SHASH model is a substantial improvement over the pilot
522 Gaussian model, as seen in the binned moments diagnostic plot, fig. 16. The fits
523 to mean and standard deviation are about as good, but only the SHASH model
524 captures the skew and kurtosis and how they vary with size.
- 525 9. Finally, a simulation study was done to see how well the shrinkage approach re-
526 covers known year effects. Fig. 17 illustrates the results for the intercept coefficient.
527 Panels A,B,C,D are the fixed-effects estimates, recommended shrunk estimates, more
528 strongly shrunk estimates, and estimates from lmer using the fitted standard devia-
529 tion from the pilot Gaussian model. Except for lmer they all look very similar, and
530 they are. The recommended shrunk estimates had the lowest mean-square error for
531 both slope and intercept (averaged across all years), and their sample variance was
532 closest to the across-year variance in the data-generating model, but the improve-
533 ment over the fixed-effects estimates was small:

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

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

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

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

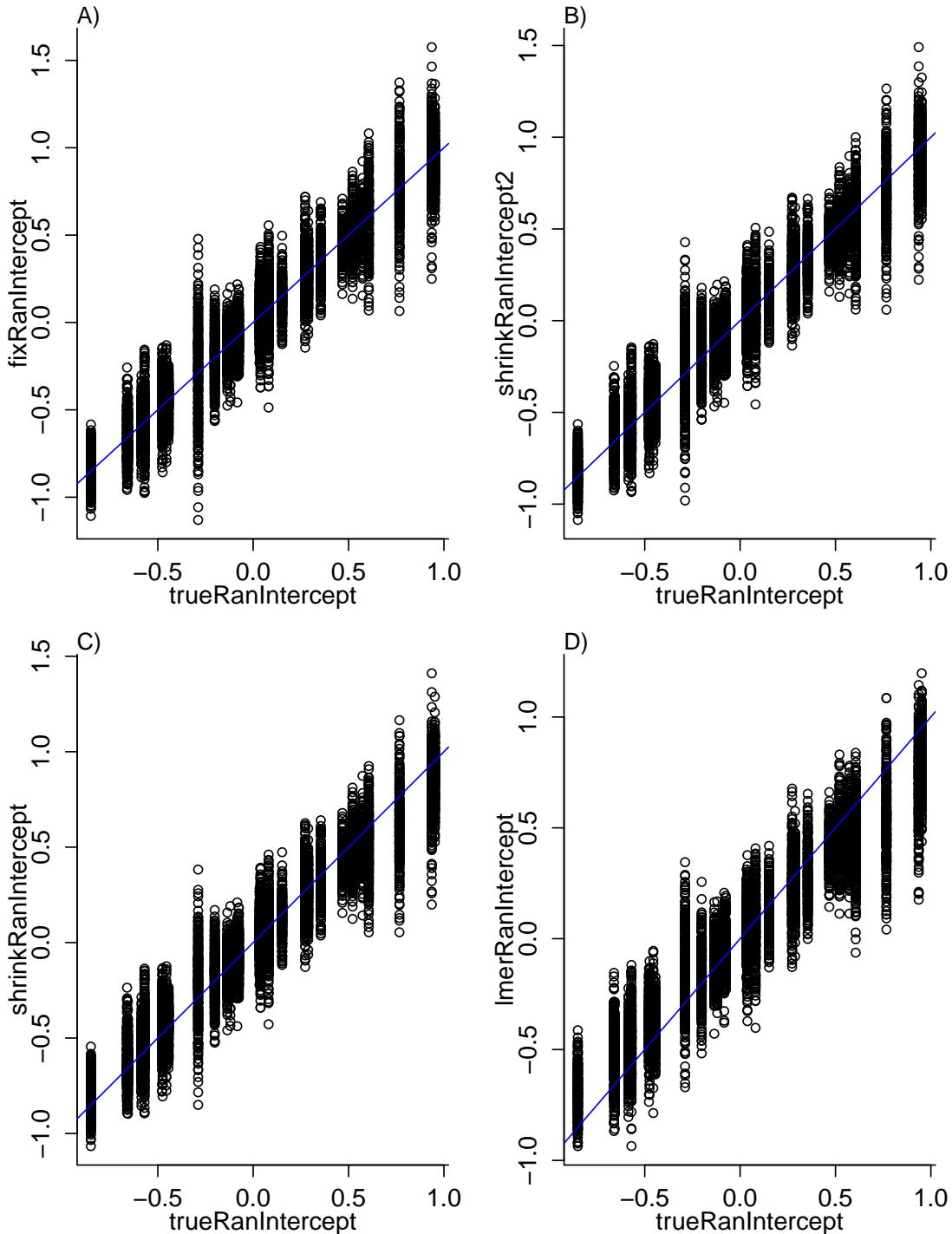


Figure 17: Comparison of “true” and estimated intercept year effects for *P. spicata*.

545 4 Case study: creosote bush, *Larrea tridentata*

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

559 4.1 A pilot Gaussian model

560 5 Discussion

561 Here are some of the issues to be discussed.

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

568 Acknowledgements

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

570 **6 Authorship statement**

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

573 **7 Data accessibility statement**

574 No original data appear in this paper. Should the paper be accepted, all computer scripts
575 supporting the results will be archived in a Zenodo package, with the DOI included at
576 the end of the article. During peer review, our data and code are available at https://github.com/texmiller/IPM_size_transitions.
577

578 **Literature Cited**

- 579 Adler, P. B., Ellner, S. P., and Levine, J. M. (2010). Coexistence of perennial plants: an
580 embarrassment of niches. *Ecology Letters*, 13(8):1019–1029.
- 581 Adler, P. B., Kleinhesselink, A., Hooker, G., Taylor, J. B., Teller, B., and Ellner, S. P.
582 (2018). Weak interspecific interactions in a sagebrush steppe? Conflicting evidence
583 from observations and experiments. *Ecology*, 99(7):1621–1632.
- 584 Anscombe, F. J. and Glynn, W. J. (1983). Distribution of the kurtosis statistic b_2 for
585 normal samples. *Biometrika*, 70(1):227–234.
- 586 Bates, D., Sarkar, D., Bates, M. D., and Matrix, L. (2007). The lme4 package. *R package
587 version*, 2(1):74.
- 588 Bruno, J. F., Ellner, S. P., Vu, I., Kim, K., and Harvell, C. D. (2011). Impacts of aspergillosis
589 on sea fan coral demography: modeling a moving target. *Ecological Monographs*,
590 81(1):123–139.
- 591 Chu, C. and Adler, P. B. (2014). When should plant population models include age
592 structure? *Journal of Ecology*, 102(2):531–543.
- 593 Chu, C. and Adler, P. B. (2015). Large niche differences emerge at the recruitment stage
594 to stabilize grassland coexistence. *Ecological Monographs*, 85(3):373–392.
- 595 Compagnoni, A., Bibian, A. J., Ochocki, B. M., Rogers, H. S., Schultz, E. L., Sneck, M. E.,
596 Elderd, B. D., Iler, A. M., Inouye, D. W., Jacquemyn, H., et al. (2016). The effect of
597 demographic correlations on the stochastic population dynamics of perennial plants.
598 *Ecological Monographs*, 86(4):480–494.
- 599 Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide
600 to bayesian model checking for ecologists. *Ecological Monographs*, 88(4):526–542.
- 601 Cooch, E. G. and White, G. C. (2020, accessed 5/17/2020). *Program MARK - a 'gentle
602 introduction'*. Available at phidot.org.
- 603 Coulson, T. (2012). Integral projections models, their construction and use in posing
604 hypotheses in ecology. *Oikos*, 121(9):1337–1350.
- 605 Crone, E. E. (2016). Contrasting effects of spatial heterogeneity and environmental
606 stochasticity on population dynamics of a perennial wildflower. *Journal of Ecology*,
607 104(2):281–291.

- 608 Czachura, K. and Miller, T. E. (2020). Demographic back-casting reveals that subtle
609 dimensions of climate change have strong effects on population viability. *Journal of*
610 *Ecology*.
- 611 D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g1.
612 *Biometrika*, pages 679–681.
- 613 Davis, C. (2015). *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.
- 614 Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying
615 a new structured population model. *Ecology*, 81(3):694–708.
- 616 Elderd, B. D. and Miller, T. E. (2016). Quantifying demographic uncertainty: Bayesian
617 methods for integral projection models. *Ecological Monographs*, 86(1):125–144.
- 618 Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E., and Rees, M. (2022).
619 A critical comparison of integral projection and matrix projection models for demo-
620 graphic analysis: Comment. *Ecology*, 103(10):e3605.
- 621 Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modeling of Structured Popula-*
622 *tions: A Practical Guide to the Integral Projection Model*. Springer, New York.
- 623 Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in
624 animal populations. *Ecology*, 79:2531 – 2538.
- 625 Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New
626 York, 2 edition.
- 627 Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed
628 models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.
- 629 Hérault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C. T., Wagner,
630 F., and Baraloto, C. (2011). Functional traits shape ontogenetic growth trajectories of
631 rain forest tree species. *Journal of ecology*, 99(6):1431–1440.
- 632 Jones, M. and Pewsey, A. (2009). *Biometrika*, 96:761 – 780.
- 633 Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis.
634 *The American Statistician*, 65(2):89 – 95.
- 635 Komsta, L. and Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and
636 related tests. *R package version*, 14(1).

- 637 Link, W. A. and Nichols, J. D. (1994). On the importance of sampling variance to inves-
638 tigations of temporal variation in animal population size. *Oikos*, 69(3):539 – 544.
- 639 Louthan, A. M., Keighron, M., Kiekebusch, E., Cayton, H., Terando, A., and Morris, W. F.
640 (2022). Climate change weakens the impact of disturbance interval on the growth rate
641 of natural populations of venus flytrap. *Ecological Monographs*, 92(4):e1528.
- 642 McGillivray, H. (1986). Skewness and asymmetry: measures and orderings. *Annals of
643 Statistics*, 14:994–1011.
- 644 Metcalf, C. J. E., Ellner, S. P., Childs, D. Z., Salguero-Gómez, R., Merow, C., McMahon,
645 S. M., Jongejans, E., and Rees, M. (2015). Statistical modelling of annual variation for
646 inference on stochastic population dynamics using Integral Projection Models. *Methods
647 in Ecology and Evolution*, 6:1007–1017.
- 648 Miller, T. E. (2020). Long-term study of tree cholla demography in the los pinos
649 mountains, sevilleta national wildlife refuge. [https://doi.org/10.6073/pasta/
650 dd06df3f950afe4a4642306182237d13](https://doi.org/10.6073/pasta/dd06df3f950afe4a4642306182237d13).
- 651 Miller, T. E., Louda, S. M., Rose, K. A., and Eckberg, J. O. (2009). Impacts of insect
652 herbivory on cactus population dynamics: experimental demography across an envi-
653 ronmental gradient. *Ecological Monographs*, 79(1):155–172.
- 654 Ohm, J. R. and Miller, T. E. (2014). Balancing anti-herbivore benefits and anti-pollinator
655 costs of defensive mutualists. *Ecology*, 95(10):2924–2935.
- 656 Ozgul, A., Childs, D. Z., Oli, M. K., Armitage, K. B., Blumstein, D. T., Olson, L. E., Tul-
657 japurkar, S., and Coulson, T. (2010). Coupled dynamics of body mass and population
658 growth in response to environmental change. *Nature*, 466(7305):482–485.
- 659 Peterson, M. L., Morris, W., Linares, C., and Doak, D. (2019). Improving structured
660 population models with more realistic representations of non-normal growth. *Methods
661 in Ecology and Evolution*, 10(9):1431–1444.
- 662 Plard, F., Schindler, S., Arlettaz, R., and Schaub, M. (2018). Sex-specific heterogene-
663 ity in fixed morphological traits influences individual fitness in a monogamous bird
664 population. *The American Naturalist*, 191(1):106–119.
- 665 Rees, M., Childs, D. Z., and Ellner, S. P. (2014). Building integral projection models: a
666 user's guide. *Journal of Animal Ecology*, 83(3):528–545.

- 667 Salguero-Gómez, R. and Casper, B. B. (2010). Keeping plant shrinkage in the demo-
668 graphic loop. *Journal of Ecology*, 98(2):312–323.
- 669 Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky,
670 C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020).
671 Robustness of linear mixed-effects models to violations of distributional assumptions.
672 *Methods in ecology and evolution*, 11(9):1141–1152.
- 673 Schultz, E. L., Eckberg, J. O., Berg, S. S., Louda, S. M., and Miller, T. E. (2017). Native
674 insect herbivory overwhelms context dependence to limit complex invasion dynamics
675 of exotic weeds. *Ecology letters*, 20(11):1374–1384.
- 676 Snyder, R. E. and Ellner, S. P. (2018). Pluck or luck: Does trait variation or chance drive
677 variation in lifetime reproductive success? *American Naturalist*, 191(4):E90 – E107.
- 678 Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location
679 scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46.
- 680 Stubberud, M. W., Vindenes, Y., Vøllestad, L. A., Winfield, I. J., Stenseth, N. C., and Lan-
681 gangen, Ø. (2019). Effects of size-and sex-selective harvesting: An integral projection
682 model approach. *Ecology and Evolution*, 9(22):12556–12570.
- 683 Teller, B. J., Adler, P. B., Edwards, C. B., Hooker, G., and Ellner, S. P. (2016). Linking
684 demography with drivers: climate and competition. *Methods in Ecology and Evolution*,
685 7(2):171–183.
- 686 Tredennick, A. T., Teller, B. J., Adler, P. B., Hooker, G., and Ellner, S. P. (2018). Size-by-
687 environment interactions: a neglected dimension of species' responses to environmen-
688 tal variation. *Ecology Letters*, 21(12):1757–1770.
- 689 Wan, X., Wang, W., Liu, J., and Tong, T. (2014). Estimating the sample mean and stan-
690 dard deviation from the sample size, median, range and/or interquartile range. *BMC*
691 *medical research methodology*, 14:1–13.
- 692 Williams, J. L., Miller, T. E., and Ellner, S. P. (2012). Avoiding unintentional eviction from
693 integral projection models. *Ecology*, 93(9):2008–2014.
- 694 Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and
695 Hall/CRC, 2 edition.
- 696 Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational*
697 *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

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

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

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

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

737 Thus, the two parameterizations are related by

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

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

746 S.2 Estimating mixed-effects models using shrinkage

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

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

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

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

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

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

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

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

787 If there are T years, the matrix `U` specified in this way has $2T$ columns corresponding to
788 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:

```

792 LogLik=function(pars,new.size,U){
793   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
794   mu = U%*%pars1;
795   sigma = exp(pars2[1]+pars2[2]*mu);
796   dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)
797 }
```

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

```

800 start=c(runif(ncol(U)), rep(0,3))
801 out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,
802   method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);
803 coefs = out$estimate; # parameters
804 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

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

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

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

$$829 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

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

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

$$837 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

838 $\hat{\sigma}^2$ estimates the variance of the distribution from which the a_t are assumed to be drawn.

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

$$843 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

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

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

$$851 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (S.10)$$

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

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

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

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

```
872 # Variance-covariance matrices for intercepts and slopes
873 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
874 # Extract year-specific intercepts, center them to zero
875 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
876
877 # Estimate sigma^2
878 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
879           (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
880
881 # Shrink deviations from the mean
```

```

882 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
883
884 # Do it all again for the slopes
885 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
886 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
887 (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
888 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

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

