

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, using software tools available
96 now.¹ Like any recipe, users may need to make substitutions or add ingredients to
97 suit their situation. Our approach emphasizes graphical diagnostics for developing and

¹Our statements about what is available now are based on what tools reliably deliver in our experience, not on what they promise.

98 evaluating growth models, rather than a process centered on statistical model selection.
99 Through a set of empirical case studies we demonstrate how a simple workflow, using
100 tools that were nonexistent or not readily available when IPMs first came into use, makes
101 it straightforward and relatively easy to identify when the default model is a poor fit to
102 the data, and to then choose and fit a substantially better growth model that is no harder
103 to use in practice. We illustrate our approach by revisiting published IPM analyses that
104 assumed Gaussian growth, including our own previous work. In each case, the Gaussian
105 assumption does not stand up to close scrutiny. We illustrate how we could have done
106 better, and the consequences of “doing better” for our ecological inferences. All of our
107 analyses may be reproduced from code and data that are publicly available (see Data
108 accessibility statement).

109 A workflow for growth modeling

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

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

113 This step is familiar to most IPM users, as it is the start and end of the traditional
114 workflow. A well-fitted Gaussian model accurately describes the mean and variance
115 of future size conditional on current size and possibly on other measured covari-
116 ates or random effects. This step may include model selection to identify which
117 treatment effects or environmental drivers affect the mean and/or variance of future
118 size. Non-constant variance is often fitted in a two-stage process, first fitting mean
119 growth assuming constant variance, then doing a regression relating the squared
120 residuals to initial size or the fitted mean of subsequent size. Fitting mean and
121 variance simultaneously, as can be done with R packages **mrgcv** and **nmle**, is ad-
122 vantageous when possible because incorrectly assuming constant variance can affect
123 model selection for the mean. But two-step fitting may be convenient when there
124 are multiple fixed and random effects that can affect growth variance, because the
125 fitted mean value implicitly accounts for all of them. We illustrate both one-step and
126 two-step approaches in the case studies below.

127 Allowing non-constant variance removes the need for transforming the data to
128 stabilize the growth variance. Transformation remains an option when it does not
129 create new problems (see Discussion), and it may have advantages besides variance

130 stabilization. In particular log-transformation is often appropriate for size data (Ell-
131 ner et al., 2016), and it helps to avoid eviction at small sizes.

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

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

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

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

164 In models with multiple covariates and/or random effects, each potentially affecting

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 any evident defects in the Gaussian model. As we demonstrate below, the functional forms for the mean and standard deviation (or location and scale parameters) can often 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 if 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 initial size or expected future size, rather than examining the aggregate 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).

197 respectively, of the distribution:

$$198 \quad \text{Skewness} = \frac{m_3}{\sigma^3}, \quad \text{Excess kurtosis} = \frac{m_4}{\sigma^4} - 3 \quad (1)$$

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

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

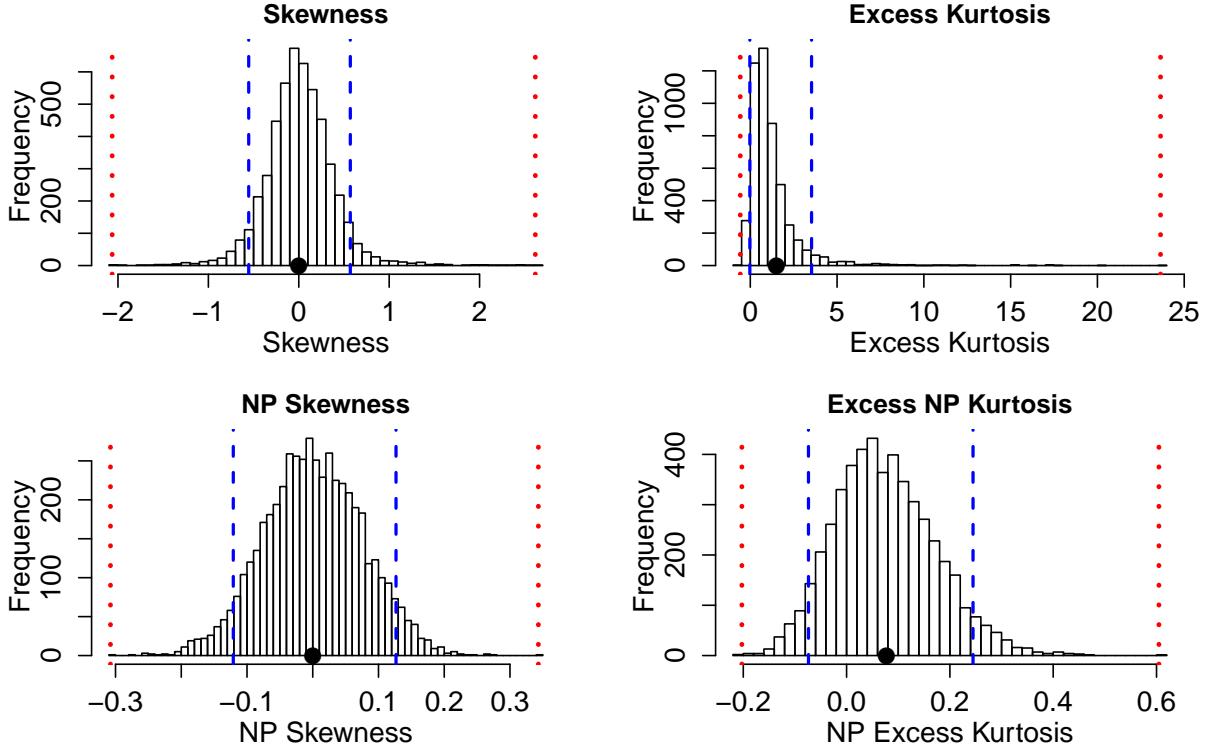


Figure 2: Histograms of skewness and kurtosis estimates using moment-based definitions (top two panels), compared with the nonparametric measures based on quantiles (bottom two panels). Note the very large differences in scale. Histograms are based on 5000 replicate draws of a sample of 200 independent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

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

We therefore use nonparametric (NP) measures of skew and kurtosis that are based on quantiles and thus are less sensitive to a few extreme 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)

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

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

$$220 \quad \text{NP Skewness} = \frac{\tau_U - \tau_L}{(\tau_U + \tau_L)}. \quad (3)$$

221 An NP Skewness of ± 0.2 says that the difference in tail sizes is 20% of their total. The
 222 range of possible values is -1 to 1. Both $\alpha = 0.25$ (sometimes called “Kelly’s skewness”) and
 223 $\alpha = 0.1$ (“Bowley’s skewness”) are common choices. We used $\alpha = 0.1$, unless
 224 otherwise stated.

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

$$226 \quad \text{NP Kurtosis} = \frac{q_{1-\alpha} - q_\alpha}{q_{0.75} - q_{0.25}}. \quad (4)$$

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

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

240 Quantile-based estimation of skewness and kurtosis carries the added value that
 241 quantile regression methods may be used to derive these properties of size transitions
 242 as continuous functions of initial size or expected future size. In the examples below,
 243 we sometimes use the **qgam** package to fit smooth additive quantile regression models,
 244 which have the flexibility to accommodate nonlinear size-dependence in skewness and
 245 kurtosis. One risk of a gam-based approach is that fitted quantiles may be too “wiggly”
 246 without constraints on their complexity. In the examples below, we limit complexity by
 247 fitting splines with $k = 4$ or $k = 6$ basis functions. For the gam-averse, other quantile
 248 regression models may be equally suitable, and we illustrate those, too. For consistency

249 with nonparametric skewness and kurtosis, in comparisons of real and simulated data
 250 below, we use quantile-based measures of location and scale, and use quantile regression
 251 to visualize these as functions of size. Specifically, following Wan et al. (2014),

$$252 \quad \text{NP Mean} = \frac{q_{0.25} + q_{0.5} + q_{0.75}}{3} \quad (5)$$

253 and

$$254 \quad \text{NP SD} = \frac{q_{0.75} - q_{0.25}}{1.35}. \quad (6)$$

255 1 Case study: lichen, *Vulpicida pinastri*

256 We begin with a simple example where current size is the only predictor of future size.
 257 Growth data for the epiphytic lichen *Vulpicida pinastri* were first analyzed by Shriner et
 258 al. 2012 and analyzed again by Peterson et al. 2019 in their study of negatively skewed
 259 growth distributions. We therefore had an *a priori* expectation of deviation from normal-
 260 ity. The authors of the original study used a mixture distribution that separated “normal
 261 growth or shrinkage” from “extreme shrinkage”. We aimed to fit a single, flexible growth
 262 model that could realistically accommodate both types of size transition without requir-
 263 ing *ad hoc* decisions about which observations of shrinkage were “extreme” or not. The
 264 data set includes 1,542 inter-annual transitions in thallus area (cm^2) observed from 2004
 265 to 2009 in Kennicott Valley, AK.

266 With initial size as the only predictor, a simple way to fit a Gaussian model with
 267 nonconstant variance is the `gam` function in `mgcv` library (Wood, 2017) using the `gaulss`
 268 family. Following a bit of model selection, we fit the mean and standard deviation of
 269 future size as second-order polynomials of current size, then derived the scaled residuals
 270 from the fitted mean and sd:

```
271 # XH is the data frame
272 # t0 and t1 are initial and final thallus area, respectively
273 fitGAU <- gam(list(t1 ~ t0 + I(t0^2), ~ t0 + I(t0^2)), data=XH, gamma=1.4, family=gaulss)
274 XH$fitted_mean = predict(fitGAU, type="response")[,1]
275 XH$fitted_sd <- 1/predict(fitGAU, type="response")[,2]
276 XH$scaledResids=residuals(fitGAU, type="response")/XH$fitted_sd
```

277 Quantile regression on the scaled residuals generates the diagnostics shown in Fig. 3
 278 (see script `Vulpicida_IPMS.R`). As expected based on previous analyses, visual analysis
 279 of the standardized residuals indicated negative skew, especially at larger sizes (Fig. 3B).
 280 We also find positive excess kurtosis for all sizes.

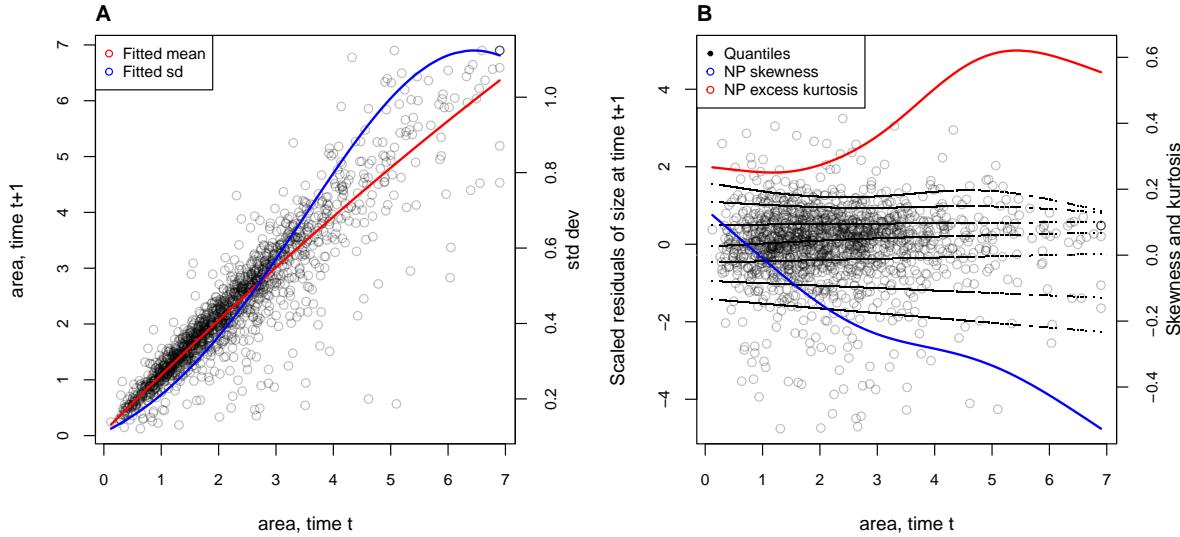


Figure 3: **A**, Size transition data for lichens, *Vulpicida pinastri*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (blue)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

Lacking clear evidence of size-dependence in kurtosis, we turned to the Johnson's *S-U* (JSU) distribution for improvement. The JSU is a four-parameter, leptokurtic distribution that can accommodate positive or negative skew; it also has the convenient property that parameters `mu` and `sigma` are the mean and standard deviation, respectively, which facilitates a natural correspondence to the pilot Gaussian model. The JSU is not available as a distribution family in any of the standard linear or additive modeling packages, to our knowledge, but that need not be a barrier for this or any other distribution as long as we can write a likelihood function (`dJSU()` is provided by `gamlss`). Following the best-fit Gaussian model, we defined `mu` and `sigma` of the JSU as second-order polynomials of initial size and, based on signals of skewness and kurtosis in the standardized residuals (Fig. 3B), we define parameter `nu` (which controls skewness) as a linear function of size and `tau` (which controls kurtosis) as a positive constant; the likelihood function therefore has nine parameters to estimate. We fit the model using the `maxLik` package and starting values for `mu` and `sigma` based on estimates from the pilot Gaussian model:

```

296  ## define function that returns the JSU negative log-likelihood
297  LogLikJSU=function(pars){
298      dJSU(size_t1,

```

```

299     mu=pars[1]+pars[2]*size_t+pars[3]*size_t^2,
300     sigma=exp(pars[4]+pars[5]*size_t+pars[6]*size_t^2),
301     nu = pars[7]+pars[8]*size_t,
302     tau = exp(pars[9]), log=TRUE)
303 }
304 ## starting parameter values
305 p0<-c(coef(fitGAU22)[1:6],0,0,0)
306 ## fit with maxlik
307 outJSU=maxLik(logLik=LogLikJSU,start=p0*exp(0.2*rnorm(length(p0))),
308 method="BHHH",control=list(iterlim=5000,printLevel=2),finalHessian=FALSE);

```

309 Data simulation from the fitted JSU model indicates a compelling improvement over the
 310 best Gaussian model, particularly in skewness and kurtosis (Fig. 4).

311 To understand the practical consequences of improved growth modeling, we as-
 312 sembled the remainder of the lichen IPM following Shriver et al. 2012. The asymptotic
 313 population growth rate based on Gaussian growth ($\lambda_{GAU} = 1.001$) differs from the JSU
 314 growth model ($\lambda_{JSU} = 0.991$) by about 1% annual population growth, in line with re-
 315 sults of Peterson et al. 2019. However, even this modest difference can lead to strongly
 316 biased estimates of extinction risk from the Gaussian model, particularly over longer
 317 time horizons (Fig. 5). We also explored differences in other life history metrics (Table).²
 318 For example, the JSU growth model predicts values for mean lifespan, mean lifetime
 319 reproductive success, and mean age at reproduction that are 19%, 25%, and 14% lower
 320 than the Gaussian growth model. In this case study, properly modeling non-normal size
 321 transitions – which was easy to do with a few extra lines of code – can have important
 322 effects on ecological inferences.

323 One could argue that the lichen data set was a convenient “straw man” to disqualify
 324 Gaussian growth, since it was recognized by the original and subsequent IPM analysts
 325 that this species requires a skewed distribution of size transitions (Peterson et al., 2019;
 326 Shriver et al., 2012). In all remaining case studies, including those in the Appendix,
 327 we re-examine growth data that were modeled as Gaussian by the data originators in
 328 published IPM studies.

²What is the best way to cite Chrissy Hernandez' life history functions?

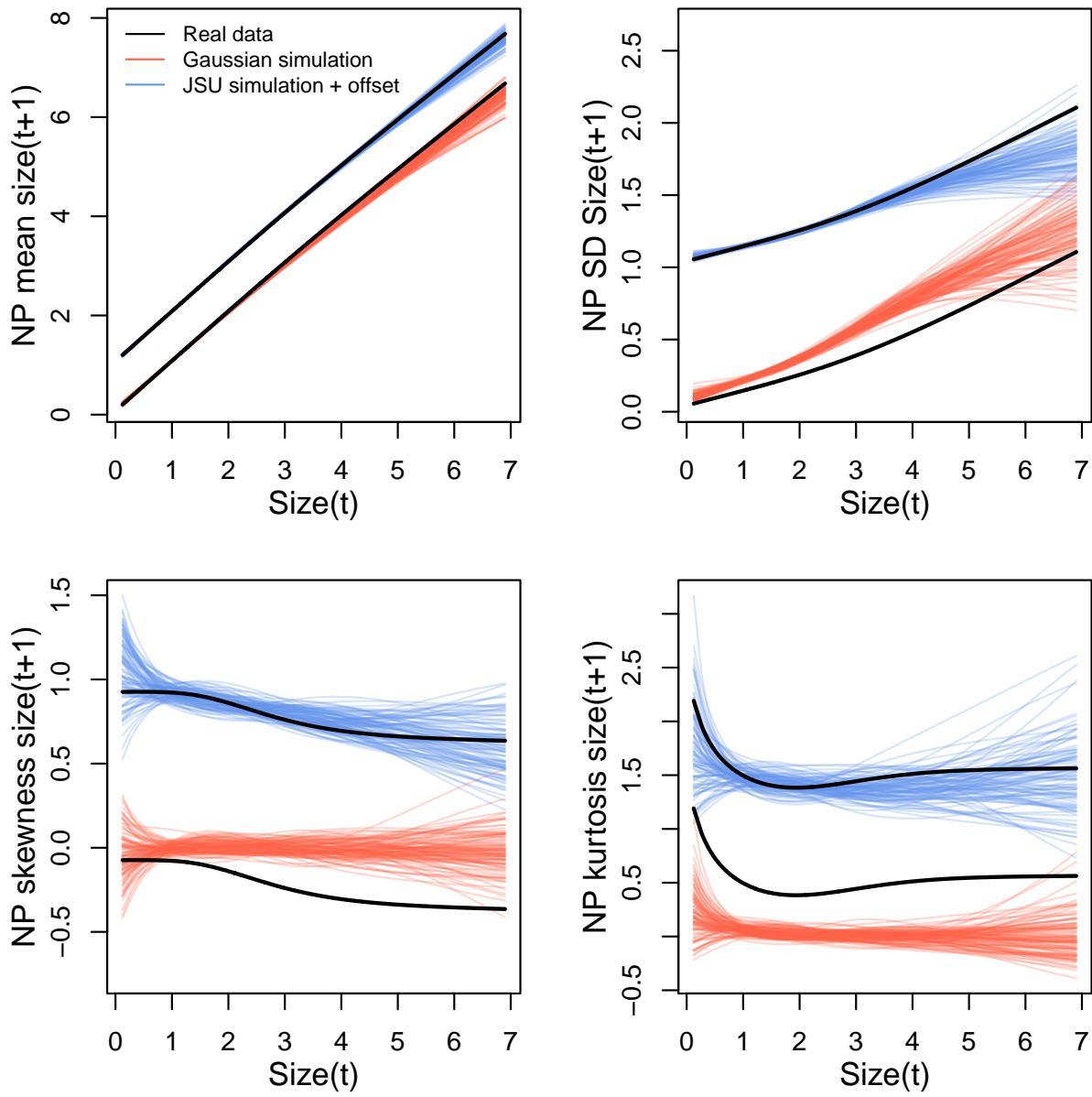


Figure 4: Comparisons among real lichen data and data simulated from Gaussian and JSU growth models for NP mean, NP standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Figure made by script `Vuplicida_IPMs.R`.

329 2 Case study: tree cholla cactus, *Cylindriopuntia imbricata*

330 The next case study, focusing on the tree cholla cactus *Cylindriopuntia imbricata* at the
 331 Sevilleta Long-Term Ecological Research site in central New Mexico, adds a new feature
 332 on top of the simple size-dependent regressions in the previous study: random effects

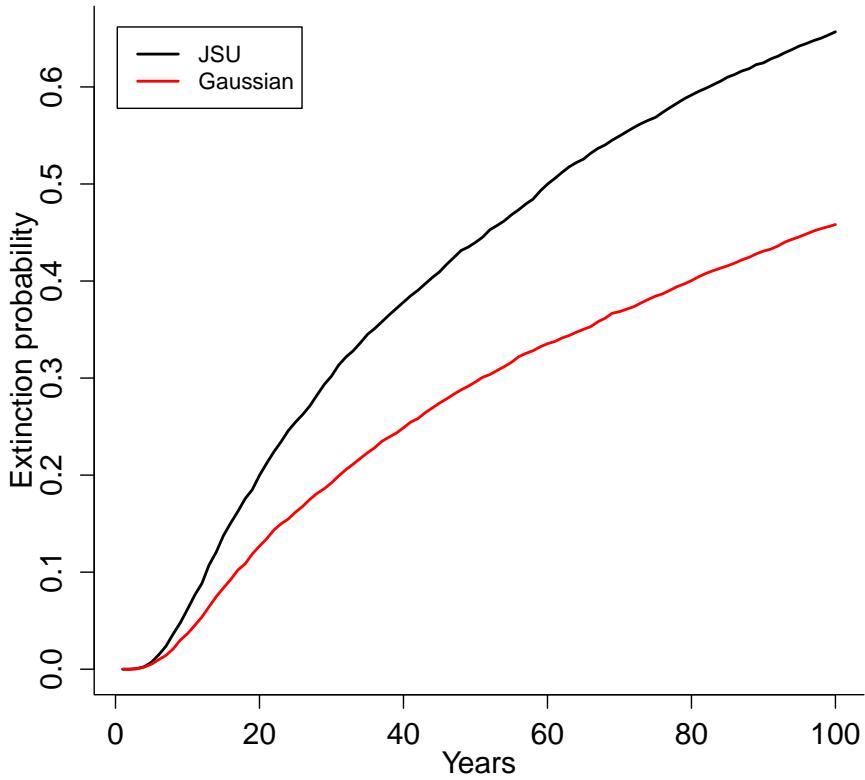


Figure 5: Extinction risk estimated from individual-based simulation of IPMs based on Gaussian and Johnson's S-U (JSU) growth distributions. Figure made by script Vuplicida_IPMs.R.

associated with temporal (year) and spatial (plot) environmental heterogeneity. This long-term study of cactus demography was initiated in 2004 and different subsets of the data have been analyzed in various IPM studies, all using Gaussian growth kernels (Compagnoni et al., 2016; Czachura and Miller, 2020; Elderd and Miller, 2016; Miller et al., 2009; Ohm and Miller, 2014). In fact, (Elderd and Miller, 2016) presented a Gaussian growth model fit to the cactus data as an example of a well fit growth function, based on a marginal distribution of residuals that appeared approximately Gaussian and posterior predictive checks (PPCs) of a Bayesian model that suggested consistency between the real data and data simulated from the fitted model (Fig. 4 in (Elderd and Miller, 2016)).

While PPCs and the associated “Bayesian P-value” are popular diagnostic tools, they are often considered to be too conservative (Conn et al., 2018; Zhang, 2014), failing

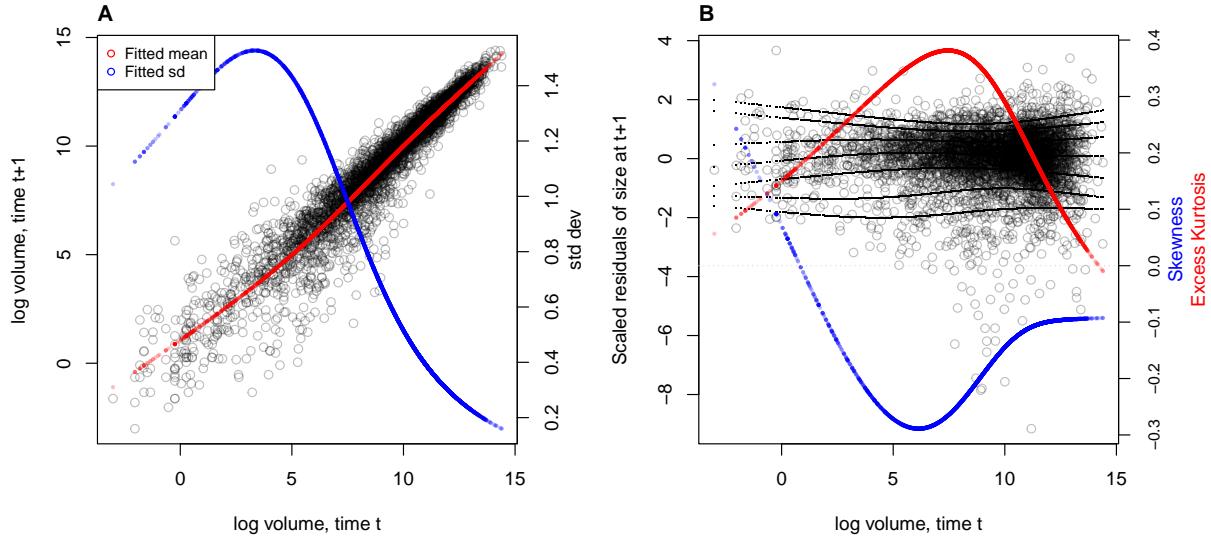


Figure 6: **A**, Size transition data for tree cholla cacti, *Cylindropuntia imbricata*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess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

345 to reject marginally bad models even though they are very effective in rejecting models
 346 that are terrible. The choice of discrepancy function (the statistic used to compare real
 347 and simulated data) can also be limiting: in our previous work, we used a discrepancy
 348 function focused on variance (the sum of the squared residuals), so we had a built-in
 349 blind-spot for mismatches in higher moments. In the clarity of hindsight, the PPC gave
 350 a false sense of security; the Gaussian was a poor choice all along.

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

356 We begin the growth modeling workflow, as above, with a generalized additive
 357 model with the mean and standard deviation of size in year $t + 1$ modeled as function
 358 of size in year t , with random intercepts for year and plot and assuming normally dis-
 359 tributed residuals (`family=gaulss()`). The standardized residuals, accounting for size-
 360 dependent residual variance (Fig. 6A), show clear signals of negative skew and positive

361 excess kurtosis across most of the size distribution but strongest in the middle of the size
362 distribution (Fig. 6B).

363 To better capture size transitions, we need a distribution with negative skew and
364 positive excess kurtosis, but both of which may be negligible at some sizes. We first tried
365 Johnson's S_U and then the skewed t distributions, both of which are limited to positive
366 excess kurtosis. Both distributions provided some improvement over the Gaussian, but
367 were not happy with the fit of either. Iterating through the workflow (Fig. 1), we arrived
368 at the SHASH distribution, which is more flexible than either the JSU or skewed t , capa-
369 ble of capturing a greater range of kurtosis for a given amount of skew, and vice versa
370 (Jones and Pewsey (2009); Appendix S.1). Furthermore, fitting the SHASH as a general-
371 ized additive model with **mgcv** allowed for flexible, non-monotonic size-dependence in
372 skewness and kurtosis without the need for model selection on specific size-dependent
373 functions; through iterations of trial and error, we found this flexibility was necessary to
374 generate simulated data that compared favorably to the real data. The other distributions
375 that we tried are not available as **mgcv** families, so we fit these with custom maximum
376 likelihood functions, an approach we illustrate in the next case study. The final growth
377 model was similar to the SHASH gam in the coral case study, but with random intercepts
378 for the location parameter, representing spatial and temporal heterogeneity:

```
379 fit_shash <- gam(list(logvol_t1 ~ s(logvol_t,k=4) +  
380   s(plot,bs="re") + s(year_t,bs="re"), # location  
381   ~ s(logvol_t,k=4), # log-scale  
382   ~ s(logvol_t,k=4), # skewness  
383   ~ s(logvol_t,k=4)), # log-kurtosis  
384   data = CYIM_grow,  
385   family = shash,  
386   optimizer = "efs")
```

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

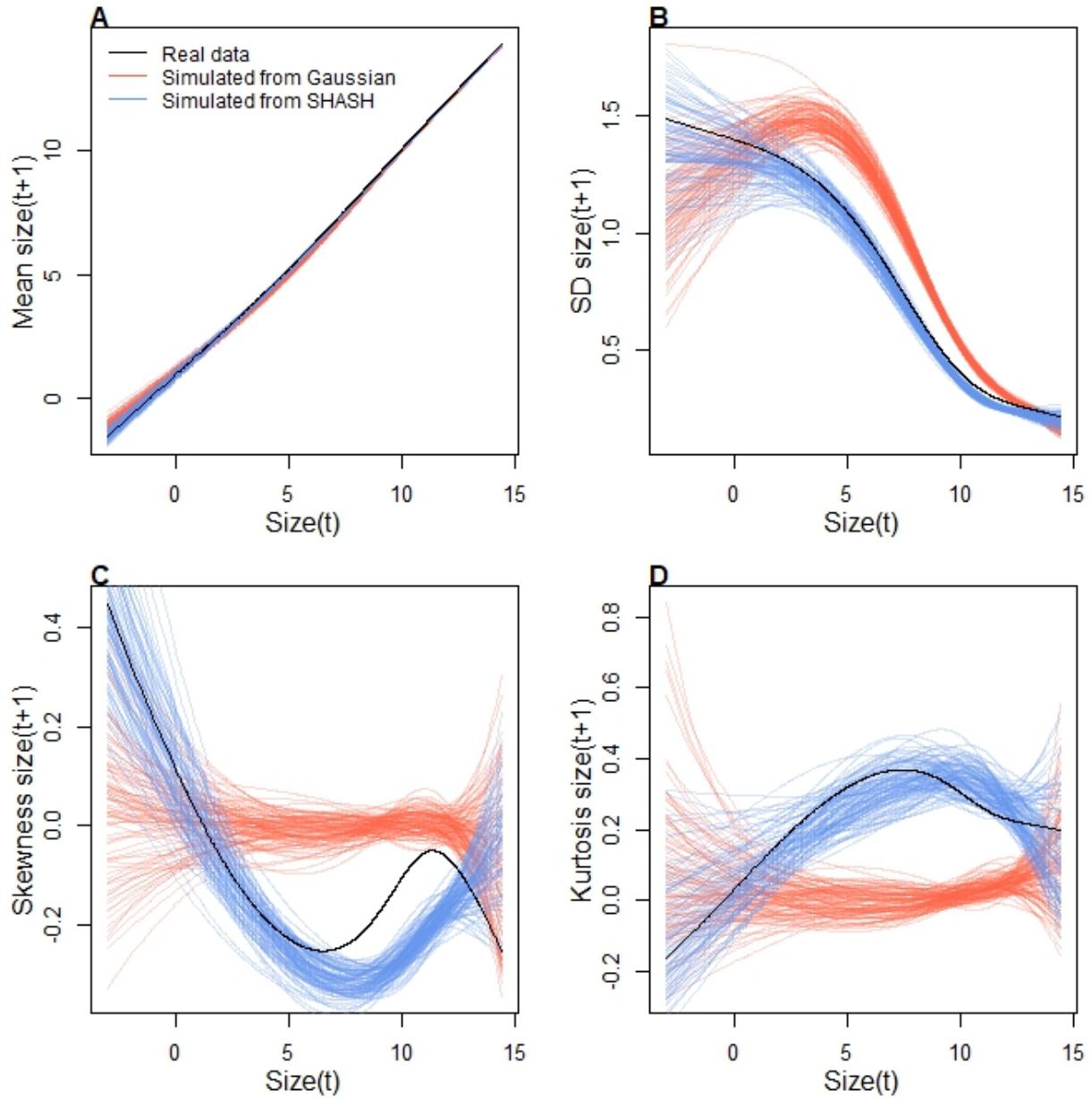


Figure 7: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

We explored how improved growth modeling influenced IPM results, leveraging the plot and year structure of the study design to quantify spatial and temporal variance in fitness. We used the fitted random effects from the vital rate models to estimate the asymptotic growth rate for each year (λ_t), centered on the average plot, and for each plot (λ_p), centered on the average year. This allowed us to quantify demographic

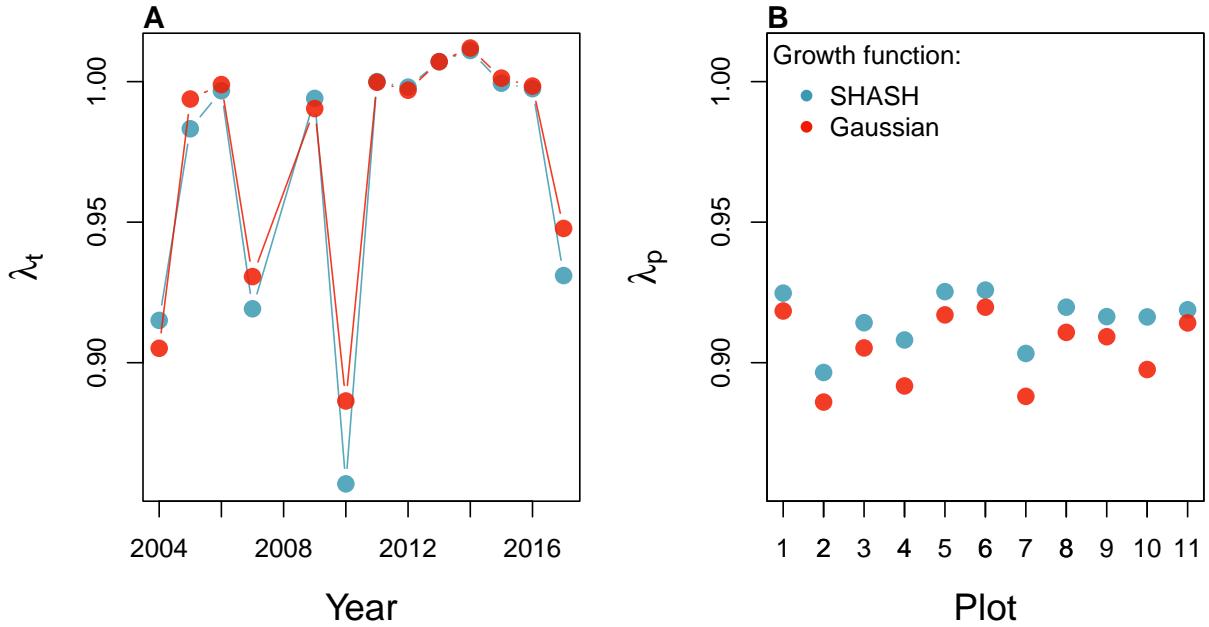


Figure 8: Temporal (A) and spatial (B) heterogeneity in fitness for the tree cholla cactus (*Cylindriopuntia imbricata*) predicted by IPMs using Gaussian or SHASH growth models. Figure made by script `cactus_growth_modeling_qgam.R`.

401 variance associated with temporal and spatial heterogeneity. We found that the Gaus-
 402 sian growth model tended to over-estimate λ_t , particularly in the harshest years (Fig.
 403 8A), and thus under-estimated temporal variance in fitness ($Var(\lambda_{t(Gaussian)}) = 0.0018$,
 404 $Var(\lambda_{t(SHASH)}) = 0.0023$). The opposite was true for plot-to-plot variation (Fig. 8B),
 405 where the Gaussian model under-estimated λ_p and over-estimated spatial variance in
 406 fitness ($Var(\lambda_{p(Gaussian)}) = 0.00015$, $Var(\lambda_{p(SHASH)}) = 0.000088$). Across both growth
 407 models, fluctuations in fitness were stronger through time than across space. The
 408 difference in temporal variance would suggest that Gaussian growth modeling would
 409 lead to over-estimation of the stochastic growth rate λ_S , since temporal variance has
 410 a negative influence on λ_S . However, this was not the case: stochastic IPMs based
 411 on Gaussian and SHASH growth models had nearly identical stochastic growth rates
 412 ($\lambda_S(Gaussian) = 0.9906$, $\lambda_S(Gaussian) = 0.9909$). This is likely because temporal fluctu-
 413 ations in vital rates, which is where the SHASH growth model would make a difference,
 414 have a weaker influence on λ_S than the temporal fluctuations in size structure that they
 415 generate (Compagnoni et al., 2016; Ellis and Crone, 2013). Thus, depending on the target
 416 of one's analysis, modeling non-Gaussian size transitions with a Gaussian growth model
 417 could bias results in either direction, or make no difference at all.

418 2.1 Case study: lady orchid, *Orchis purpurea*

419 Our final case study examines selection on life history strategies in the lady orchid *Or-*
420 *chis purpurea*. In a prior study, Miller et al. 2012 contrasted the growth trajectories from
421 year t to $t + 1$ for plants that did or did not flower in year t , as a way to quantify costs
422 of reproduction. The different growth kernels were then used in an IPM to quantify
423 evolutionarily stable life history strategies: the optimal flowering size that balances ben-
424 efits of flowering at larger sizes against the risk of dying before reaching those sizes.
425 The original study assumed a Gaussian distribution of size transitions and allowed for
426 non-constant variance with respect to initial size. Here we re-visit that analysis applying
427 our growth modeling workflow to derive improved growth kernels for flowering and
428 non-flowering orchids. We use this case study to illustrate several new elements and
429 challenges, including modeling skewness and kurtosis as functions of expected future
430 size (instead of initial size) and using distributions that are not available as **mgcv** fam-
431 ilies. In fact, to diversify our use of software and illustrate alternatives, we do not use
432 gam's for any element of this case study.

433 The data, originated by Dr. Hans Jacquemyn and used here with permission, come
434 from 368 plants in a Belgian population that was censused annually from 2003 through
435 2011 (for this reanalysis we are using data only from the “light” habitat). Size was mea-
436 sured as leaf area (cm^3) summed over all leaves, and we analyzed the natural logarithm
437 of total leaf area as the size variable of the IPM.

438 Unlike our previous case studies, here we have multiple fixed effects that may influ-
439 ence the variance of future size. In cases such as this, we recommend modeling variance
440 as a function of expected future size, rather than initial size as we did with the corals
441 and cacti. The expected (or “fitted”) values reflect the combined influence of all fixed
442 and random effects, and therefore implicitly account for multiple sources of variation in
443 the variance. While there are several convenient software packages for simultaneously
444 modeling Gaussian mean and variance as functions of independent variables (**mgcv** for
445 gam models as we saw above, **nlme** for linear models), **modeling variance as a function**
446 **of the mean is trickier because they cannot easily be fit simultaneously**³. Here we us an
447 iterative re-weighting approach – which is not elegant, but it works. For Gaussian mod-
448 els, weights w_i can be used to indicate that the observations y_i vary in their dispersion
449 around the mean. In general, the iterative steps are:

³ 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:

$$y_i = \mu_i + \epsilon_i$$

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

450 We iterated steps 2 and 3 until the weights did not change. In step 2, we modeled
 451 the standard deviation as a simple linear function of the expected value ($\log(f(\mu_i)) =$
 452 $\beta_0 + \beta_1 * \mu_i$) but other functions are possible, as is model selection among them. We
 453 did this for all candidate models and, for fair AIC comparison, we re-fit all candidate
 454 models with the same weights, estimated from the top model. The updated model
 455 selection continued to favor model 4, but now with a stronger improvement over the
 456 next-best model ($\Delta AIC = 3.0$).

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

```
464 orchid_GAU[[1]]<-lme(log_area_t1~ log_area_t,  

  465 weights=varExp(form=~ log_area_t),  

  466 random=~ 1|begin.year,data=orchid_grow,method="ML")
```

467 Model 3 (size \times flowering) was strongly favored, consistent with prior results that non-
 468 flowering plants have a growth advantage over flowering plants. Growth variance de-
 469 clined with initial size for both reproductive classes (Fig. 9A-B) and skewness and kurto-

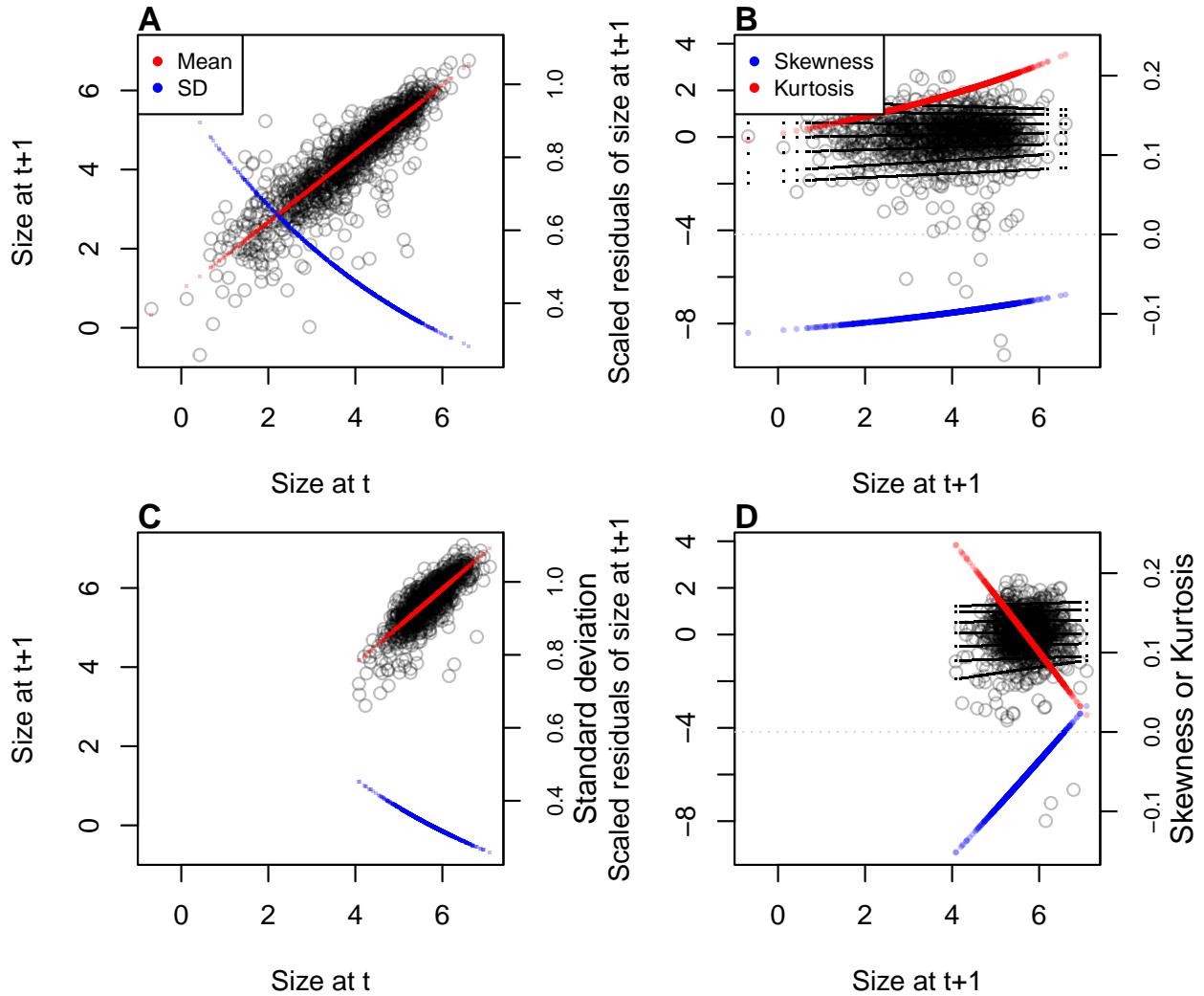


Figure 9

sis of the standardized residuals indicate strong deviations from normality (Fig. 9C-D).
 470 For most sizes, left skew and excess kurtosis were more severe for non-reproductive
 471 plants, with tail imbalance ca. 10% of their total and tail weights 10–20% fatter than
 472 Gaussian.
 473

As improvements, we explored the skewed *t* and Johnson's SU distributions, both
 474 leptokurtic distributions with flexible skewness. We were happier with the skewed *t*,
 475 which we fit in a similar way as we fit the JSU to the creosote data, setting the mean
 476 and standard deviation to the Gaussian fits and estimating free parameters controlling
 477 skewness and kurtosis:
 478

```
479 ## log_area_t1 and log_area_t are the size observations
480 ## flowering indicates reproductive status at time t (0 or 1)
```

```

481     ## GAU_fitted and GAU_sd are mean and standard deviation from lme
482     ## pars is a vector of free parameters to be estimated
483     SSTLogLik=function(pars){
484         dSST(x=log_area_t1,
485             mu=GAU_fitted,
486             sigma=GAU_sd,
487             nu = exp(pars[1] + pars[2]*log_area_t + pars[3]*as.logical(flowering) +
488             tau = exp(pars[5] + pars[6]*log_area_t + pars[7]*as.logical(flowering) -
489             log=TRUE)
490     }

```

491 `gamlss.dist:dSST` is a parameterization of the skewed t in which `mu` and `sigma` are the
492 mean and standard deviation, respectively. Based on diagnostics of the standardized
493 residuals (Fig. 9) we allowed `nu` and `tau` to vary by size and differ between flowering
494 and non-flowering plants (note that the `tau` parameter uses a $\log(x - 2)$ link function).
495 Size transition data simulated from this model corresponded favorably to the real data,
496 much better than the pilot Gaussian model, including improvements in the **standard**
497 **deviation**⁴, skewness, and kurtosis of future size (Fig. 10).

498 Finally, we used the improved growth model to revisit key results of the original
499 study. Miller et al. (2012) used the orchid IPM to estimate the evolutionarily stable strat-
500 egy (ESS) as the mean size at flowering that maximizes lifetime reproductive success
501 (R_0), given the constraint that flowering when small reduces growth and thus elevates
502 mortality risk. Repeating that analysis here, we found that improved growth modeling
503 has virtually no influence on predictions for optimal life history strategies (Fig. 11). ESS
504 flowering sizes were nearly identical between IPMs with Gaussian vs skewed t growth
505 models, and both aligned well with the observed mean flowering size (dashed vertical
506 line in Fig. 11A). Extending beyond the original study, we also explored expected re-
507 maining lifespan for different ages and sizes (R package **Rage** (Jones et al., 2022)). Gaus-
508 sian and skewed t growth models predicted nearly identical mean remaining lifespans
509 across the stage and size distribution (Fig. 11B). **However, the skewed t model predicted**
510 **consistently greater variance in remaining lifespan, nearly 10% greater at some sizes.**⁵
511 Thus, as we have seen in other case studies, the practical consequences of improved
512 growth modeling depend on what one aims to learn from the IPM.

⁴Again, the improvement here is surprising to me and I am unsure what to say about it. SPE: again, it's probably about the "usual" mean and SD versus the quantile-based NP versions.

⁵Do 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.

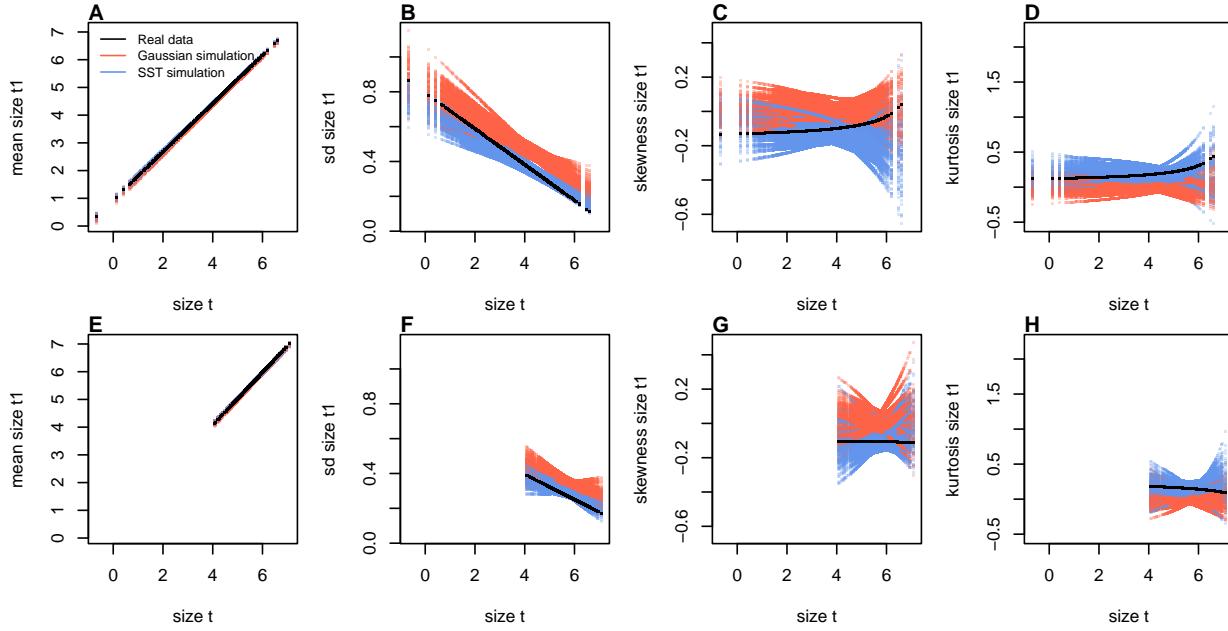


Figure 10: Comparisons between real orchid data and data simulated from Gaussian and skewed t growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (A-D) shows plants that were vegetative at the start of the transition year and bottom row (E-H) shows plants that were flowering at the start of the transition year. Figure made by script `orchid_growth_modeling_rq.R`.

513 3 Discussion

514 Much of the appeal of integral projection models has stemmed from their embrace of
 515 continuous size structure through reliance on regression-based approaches, and the po-
 516 tentially complex fixed- and random-effect structures that these approaches allow. Using
 517 familiar statistical tools and with relatively few parameters to estimate, IPM users can
 518 incorporate important sources of variation in demography and interrogate their influ-
 519 ence on ecological and evolutionary dynamics. With this opportunity comes the burden
 520 of getting it right: IPMs are good models of the populations they are intended repre-
 521 sent only insofar as the statistical models provide good fits to the underlying data. The
 522 growth sub-model is the trickiest part of “getting it right” because it defines a distri-
 523 bution of future size conditional on current size. Distributions have many properties –
 524 “moments” – and a good growth model should recapitulate the properties of real size
 525 transitions. The default assumption of normally distributed size transitions, employed
 526 overwhelmingly across 20+ years of IPM studies, is an arbitrary historical precedent.
 527 In our case studies (chosen simply because we had the data at our fingertips) and, we
 528 suspect, more broadly, skewness and excess kurtosis were common features of size tran-

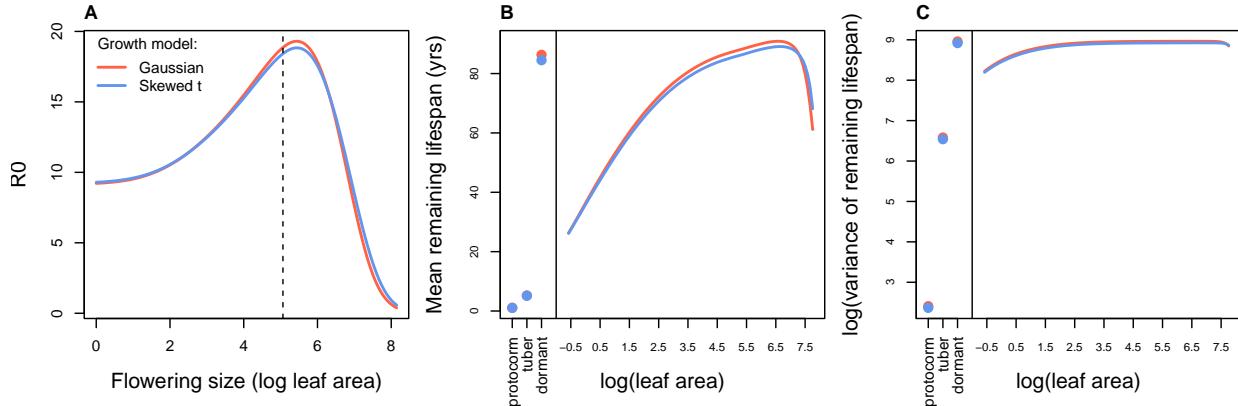


Figure 11: Orchid life history results from IPMs using Gaussian or skewed t growth models. **A**, Lifetime reproductive success (R_0) as a function of mean size of flowering. Dashed vertical line shows the observed mean flowering size. **B-C**, Mean and variance of remaining lifespan as a function of 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. [SPE: the variance of lifespan results are now correct, using Chrissy Hernandez's functions instead of RAGE. And the result is, non-Gaussian again has no effect.](#)

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

Our work follows the important contribution of Peterson et al. 2019, who were motivated by a similar problem (inadequacy of the Gaussian distribution for skewed size transitions) but arrived at different recommendations for dealing with the problem.

548 These authors developed a creative approach in which size data are transformed onto a
549 [0, 1] scale and then size transitions are modeled using beta regression. The beta distri-
550 bution can accommodate positive, negative, or zero skew; as Peterson et al. demonstrate,
551 this approach is a viable option for skewed growth data. However, beta regression also
552 has some important downsides: common beta regression packages do not fit random
553 effects(e.g., **betareg** (Cribari-Neto and Zeileis, 2010)) or do not do so reliably (in our ex-
554 perience **gamlss** is numerically unstable); the two-parameter beta distribution has rigid
555 mean-variance and skewness-kurtosis relationships that may not describe the data well;
556 and⁶. Rather than shoe-horn size transition data into a default beta distribution, we find
557 it more natural and appropriate to leverage the vast arsenal of real-valued probability
558 distributions – all of them at one’s fingertips with a few lines of code – and let the data
559 suggest appropriate alternatives based on the residuals of a pilot Gaussian model. The
560 workflow that we have envisioned requires no sacrifice of complexity in random effects
561 or covariates for the sake of modeling skewness and kurtosis.

562 While the arsenal of candidate distributions is indeed vast, in our analyses for this
563 paper, we found ourselves coming back time and again to a few usual suspects. The
564 four-parameter SHASH distribution, for example, is able to flexibly accommodate inde-
565 pendent, size-dependent variation in variance, skewness, and kurtosis, and it is available
566 as a distribution family in the well-developed **mgcv** package. In our case study analy-
567 ses it was consistently among the top non-Gaussian candidates and was our model of
568 choice for the coral, cactus, and pike data sets. While we have emphasized the im-
569 portance of moving away from a single default distribution of size transitions, if one
570 were to want or need a default distribution (e.g., in software packages for IPM con-
571 struction [cite IPMpack]) one could do worse than the SHASH. In cases where size
572 transitions are leptokurtic but consistently so across the size distribution, the Johnson’s
573 S-U (used for creosotebush) and skewed *t* (used for lady orchid) distributions were easy
574 to fit with custom likelihood functions. All of the distributions we have used (and
575 the specific parameterizations we have implemented) share the property that their lo-
576 cation and scale parameters correspond to the mean and standard deviation, which
577 is not essential but it facilitates interpretation and an intuitive connection to the pilot
578 Gaussian model. The five-parameter skewed generalized *t* (sgt) generalizes many other
579 real-valued distributions (cite) and is therefore another flexible option, but it does not
580 share the location=mean and scale=sd property, and in our experience can be hard to fit.
581 Finally, finding an appropriate non-Gaussian alternative does not solve all the problems

⁶Steve, I recall you have a beef with transformation to [0,1], so that would go here.

582 of growth modeling. “Eviction” from the approximating matrix of the IPM kernel is an
583 ever-present danger and requires vigilance to detect and correct (Williams et al., 2012).

584 In all of our case studies, non-Gaussian growth models always yielded more sat-
585 isfying fits to size transition data than the Gaussian models published in those papers.
586 However, much to our relief, none of these re-analyses yielded a “gotcha” result that
587 overturned results of the original study. In this small sampling of case studies, improved
588 growth modeling had weak to modest effects on IPM results, similar in magnitude to
589 the results of Peterson et al. (2019). We caution against taking too much comfort in this
590 outcome; we can imagine other scenarios in which the choice of the growth distribution
591 could be more consequential. It is worth noting that most of our case studies focused on
592 perennial life histories (perennial plants, corals, lichens), characterized by relatively slow
593 growth, heavy losses during recruitment, and high survival once established. Life cycles
594 such as these may be relatively robust to subtle features of the growth kernel. More
595 systematic comparative analyses may provide insight into which types of species and
596 life histories are more likely to exhibit strong skewness and kurtosis, and the conditions
597 under which demographic analysis is more or less sensitive to these features of size trans-
598 ition. It is also worth noting, as we saw in several case studies, that different outputs
599 from the same model can be more or less sensitive to the choice of growth distribution.

600 Across our case studies we have attempted to illustrate a diversity of software pack-
601 ages and computational approaches to model fitting, to reflect the diversity of prefer-
602 ences and habits that the community of IPM users bring to their own analyses. We like
603 generalized additive models (`gams`) for their flexibility and for `mgcv`’s numerous op-
604 tions for distribution families and overall speed and reliability. However, there are some
605 applications for which classical parametric regression would be preferable because re-
606 gression parameters carry biological meaning. For example, regression coefficients may
607 be targets of natural selection (Rees and Ellner, 2016) and may combine to influence
608 traits of interest such as the expected size at flowering (e.g. in Fig. 11A), a function of
609 the intercept and slope of the size-dependent flowering function (Metcalf et al., 2003).
610 Some potentially useful but relatively obscure distributions may not be available in lin-
611 ear modeling software packages, but that should not be a barrier to their use: as we have
612 illustrated in several case studies, custom likelihood functions open up diverse possi-
613 bilities for non-Gaussian growth modeling without sacrificing the complex, multi-level
614 features that one might be accustomed to fitting in `lme4`, for example. We have illus-
615 trated fitting growth models using maximum likelihood but Bayesian analysis is another
616 option that may further broaden the options of non-Gaussian candidate distributions
617 and may help estimate hard-to-fit parameters through the brute force of sampling algo-

618 rithms. Bayesian analysis also provides a natural way to propagate uncertainty from the
619 vital rate sub-models through the predictions of the IPM (Elderd and Miller, 2016).

620 This paper has focused on size transitions, but IPMs have been increasingly ex-
621 tended in ways that capture other continuous state variables, and the same problems
622 and solutions we propose should apply in those cases. For example, IPMs can be used
623 to model infectious disease dynamics, where hosts may exhibit continuous variation in
624 infection load (e.g., parasite density), and host vital rate processes depend on infection
625 load (Metcalf et al., 2016; Wilber et al., 2016). Such models must define probabilities of
626 future infection load conditional on current load, and would therefore benefit from the
627 same modeling workflow that we have outlined for size transitions.

628 **3.1 Conclusion**

629 **Acknowledgements**

630 This research was supported by US NSF grants DEB-1933497 to SPE and DEB-1754468,
631 2208857, and 2225027 to TEXM.

632 **4 Authorship statement**

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

635 **5 Data accessibility statement**

636 No original data appear in this paper. Should the paper be accepted, all computer scripts
637 supporting the results will be archived in a Zenodo package, with the DOI included at
638 the end of the article. During peer review, our data and code are available at https://github.com/texmiller/IPM_size_transitions.
639

640 **Literature Cited**

- 641 Anscombe, F. J. and Glynn, W. J. (1983). Distribution of the kurtosis statistic b_2 for
642 normal samples. *Biometrika*, 70(1):227–234.
- 643 Bates, D., Sarkar, D., Bates, M. D., and Matrix, L. (2007). The lme4 package. *R package
644 version*, 2(1):74.
- 645 Bruno, J. F., Ellner, S. P., Vu, I., Kim, K., and Harvell, C. D. (2011). Impacts of aspergillosis
646 on sea fan coral demography: modeling a moving target. *Ecological Monographs*,
647 81(1):123–139.
- 648 Compagnoni, A., Bibian, A. J., Ochocki, B. M., Rogers, H. S., Schultz, E. L., Sneck, M. E.,
649 Elderd, B. D., Iler, A. M., Inouye, D. W., Jacquemyn, H., et al. (2016). The effect of
650 demographic correlations on the stochastic population dynamics of perennial plants.
651 *Ecological Monographs*, 86(4):480–494.
- 652 Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide
653 to bayesian model checking for ecologists. *Ecological Monographs*, 88(4):526–542.
- 654 Cooch, E. G. and White, G. C. (2020, accessed 5/17/2020). *Program MARK - a 'gentle
655 introduction'*. Available at phidot.org.
- 656 Coulson, T. (2012). Integral projections models, their construction and use in posing
657 hypotheses in ecology. *Oikos*, 121(9):1337–1350.
- 658 Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in r. *Journal of statistical software*,
659 34:1–24.
- 660 Crone, E. E. (2016). Contrasting effects of spatial heterogeneity and environmental
661 stochasticity on population dynamics of a perennial wildflower. *Journal of Ecology*,
662 104(2):281–291.
- 663 Czachura, K. and Miller, T. E. (2020). Demographic back-casting reveals that subtle
664 dimensions of climate change have strong effects on population viability. *Journal of
665 Ecology*.
- 666 D'Agostino, R. B. (1970). Transformation to normality of the null distribution of g_1 .
667 *Biometrika*, pages 679–681.
- 668 Davis, C. (2015). *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.

- 669 Drees, T., Ochocki, B. M., Collins, S. L., and Miller, T. E. (2023). Demography and
670 dispersal at a grass-shrub ecotone: a spatial integral projection model for woody plant
671 encroachment. *Ecological Monographs*, page e1574.
- 672 Easterling, M. R., Ellner, S. P., and Dixon, P. M. (2000). Size-specific sensitivity: applying
673 a new structured population model. *Ecology*, 81(3):694–708.
- 674 Elderd, B. D. and Miller, T. E. (2016). Quantifying demographic uncertainty: Bayesian
675 methods for integral projection models. *Ecological Monographs*, 86(1):125–144.
- 676 Ellis, M. M. and Crone, E. E. (2013). The role of transient dynamics in stochastic popula-
677 tion growth for nine perennial plants. *Ecology*, 94(8):1681–1686.
- 678 Ellner, S. P., Adler, P. B., Childs, D. Z., Hooker, G., Miller, T. E., and Rees, M. (2022).
679 A critical comparison of integral projection and matrix projection models for demo-
680 graphic analysis: Comment. *Ecology*, 103(10):e3605.
- 681 Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modeling of Structured Popula-*
682 *tions: A Practical Guide to the Integral Projection Model*. Springer, New York.
- 683 Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in
684 animal populations. *Ecology*, 79:2531 – 2538.
- 685 Gu, C. (2013). *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New
686 York, 2 edition.
- 687 Hadfield, J. D. et al. (2010). Mcmc methods for multi-response generalized linear mixed
688 models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22.
- 689 Hérault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C. T., Wagner,
690 F., and Baraloto, C. (2011). Functional traits shape ontogenetic growth trajectories of
691 rain forest tree species. *Journal of ecology*, 99(6):1431–1440.
- 692 Jones, M. and Pewsey, A. (2009). Sinh-arcsinh distributions. *Biometrika*, 96:761 – 780.
- 693 Jones, M. C., Rosco, J. F., and Pewsey, A. (2011). Skewness-invariant measures of kurtosis.
694 *The American Statistician*, 65(2):89 – 95.
- 695 Jones, O. R., Barks, P., Stott, I., James, T. D., Levin, S., Petry, W. K., Capdevila, P., Che-
696 Castaldo, J., Jackson, J., Römer, G., et al. (2022). Rcompadre and rage—two r pack-
697 ages to facilitate the use of the comadre and comadre databases and calculation of

- 698 life-history traits from matrix population models. *Methods in Ecology and Evolution*,
699 13(4):770–781.
- 700 Komsta, L. and Novomestky, F. (2015). Moments, cumulants, skewness, kurtosis and
701 related tests. *R package version*, 14(1).
- 702 Link, W. A. and Nichols, J. D. (1994). On the importance of sampling variance to inves-
703 tigations of temporal variation in animal population size. *Oikos*, 69(3):539 – 544.
- 704 Louthan, A. M., Keighron, M., Kiekebusch, E., Cayton, H., Terando, A., and Morris, W. F.
705 (2022). Climate change weakens the impact of disturbance interval on the growth rate
706 of natural populations of venus flytrap. *Ecological Monographs*, 92(4):e1528.
- 707 McGillivray, H. (1986). Skewness and asymmetry: measures and orderings. *Annals of*
708 *Statistics*, 14:994–1011.
- 709 Metcalf, C. J. E., Ellner, S. P., Childs, D. Z., Salguero-Gómez, R., Merow, C., McMahon,
710 S. M., Jongejans, E., and Rees, M. (2015). Statistical modelling of annual variation for
711 inference on stochastic population dynamics using Integral Projection Models. *Methods*
712 *in Ecology and Evolution*, 6:1007–1017.
- 713 Metcalf, C. J. E., Graham, A. L., Martinez-Bakker, M., and Childs, D. Z. (2016). Oppor-
714 tunities and challenges of i ntegral p rojection m odels for modelling host-parasite
715 dynamics. *Journal of Animal Ecology*, 85(2):343–355.
- 716 Metcalf, J. C., Rose, K. E., and Rees, M. (2003). Evolutionary demography of monocarpic
717 perennials. *Trends in Ecology & Evolution*, 18(9):471–480.
- 718 Miller, T. E. (2020). Long-term study of tree cholla demography in the los pinos
719 mountains, sevilleta national wildlife refuge. [https://doi.org/10.6073/pasta/
720 dd06df3f950afe4a4642306182237d13](https://doi.org/10.6073/pasta/dd06df3f950afe4a4642306182237d13).
- 721 Miller, T. E., Louda, S. M., Rose, K. A., and Eckberg, J. O. (2009). Impacts of insect
722 herbivory on cactus population dynamics: experimental demography across an envi-
723 ronmental gradient. *Ecological Monographs*, 79(1):155–172.
- 724 Miller, T. E., Williams, J. L., Jongejans, E., Brys, R., and Jacquemyn, H. (2012). Evolution-
725 ary demography of iteroparous plants: incorporating non-lethal costs of reproduction
726 into integral projection models. *Proceedings of the Royal Society B: Biological Sciences*,
727 279(1739):2831–2840.

- 728 Ochocki, B. M., Drees, T., and Miller, T. E. (2023). Density-dependent demography of
729 creosote bush (*larrea tridentata*) along grass-shrub ecotones. <https://doi.org/10.6073/pasta/ca53c16f16dcf9fb11f3ee99ea5445ac>.
- 730
- 731 Ohm, J. R. and Miller, T. E. (2014). Balancing anti-herbivore benefits and anti-pollinator
732 costs of defensive mutualists. *Ecology*, 95(10):2924–2935.
- 733 Ozgul, A., Childs, D. Z., Oli, M. K., Armitage, K. B., Blumstein, D. T., Olson, L. E., Tul-
734 japurkar, S., and Coulson, T. (2010). Coupled dynamics of body mass and population
735 growth in response to environmental change. *Nature*, 466(7305):482–485.
- 736 Peterson, M. L., Morris, W., Linares, C., and Doak, D. (2019). Improving structured
737 population models with more realistic representations of non-normal growth. *Methods
738 in Ecology and Evolution*, 10(9):1431–1444.
- 739 Plard, F., Schindler, S., Arlettaz, R., and Schaub, M. (2018). Sex-specific heterogeneity
740 in fixed morphological traits influences individual fitness in a monogamous bird
741 population. *The American Naturalist*, 191(1):106–119.
- 742 Rees, M., Childs, D. Z., and Ellner, S. P. (2014). Building integral projection models: a
743 user's guide. *Journal of Animal Ecology*, 83(3):528–545.
- 744 Rees, M. and Ellner, S. P. (2016). Evolving integral projection models: evolutionary
745 demography meets eco-evolutionary dynamics. *Methods in Ecology and Evolution*,
746 7(2):157–170.
- 747 Rigby, R. A., Stasinopoulos, M. D., Heller, G. Z., and De Bastiani, F. (2019). *Distributions
748 for modeling location, scale, and shape: Using GAMLS in R*. CRC press.
- 749 Salguero-Gómez, R. and Casper, B. B. (2010). Keeping plant shrinkage in the demo-
750 graphic loop. *Journal of Ecology*, 98(2):312–323.
- 751 Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky,
752 C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020).
753 Robustness of linear mixed-effects models to violations of distributional assumptions.
754 *Methods in ecology and evolution*, 11(9):1141–1152.
- 755 Schultz, E. L., Eckberg, J. O., Berg, S. S., Louda, S. M., and Miller, T. E. (2017). Native
756 insect herbivory overwhelms context dependence to limit complex invasion dynamics
757 of exotic weeds. *Ecology letters*, 20(11):1374–1384.

- 758 Shriner, R. K., Cutler, K., and Doak, D. F. (2012). Comparative demography of an epi-
759 phytic lichen: support for general life history patterns and solutions to common prob-
760 lems in demographic parameter estimation. *Oecologia*, 170:137–146.
- 761 Stasinopoulos, D. M., Rigby, R. A., et al. (2007). Generalized additive models for location
762 scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46.
- 763 Stubberud, M. W., Vindenes, Y., Vøllestad, L. A., Winfield, I. J., Stenseth, N. C., and Lan-
764 gangen, Ø. (2019). Effects of size-and sex-selective harvesting: An integral projection
765 model approach. *Ecology and Evolution*, 9(22):12556–12570.
- 766 Wan, X., Wang, W., Liu, J., and Tong, T. (2014). Estimating the sample mean and stan-
767 dard deviation from the sample size, median, range and/or interquartile range. *BMC*
768 *medical research methodology*, 14:1–13.
- 769 Wilber, M. Q., Langwig, K. E., Kilpatrick, A. M., McCallum, H. I., and Briggs, C. J. (2016).
770 Integral projection models for host–parasite systems with an application to amphibian
771 chytrid fungus. *Methods in Ecology and Evolution*, 7(10):1182–1194.
- 772 Williams, J. L., Miller, T. E., and Ellner, S. P. (2012). Avoiding unintentional eviction from
773 integral projection models. *Ecology*, 93(9):2008–2014.
- 774 Wood, S. (2017). *Generalized Additive Models: An Introduction with R*. Chapman and
775 Hall/CRC, 2 edition.
- 776 Zhang, J. L. (2014). Comparative investigation of three bayesian p values. *Computational*
777 *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 to define a centered and scaled version in which μ and σ are the mean and standard deviation.

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

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

816
$$X_{\lambda, \tau} = \sinh(e^{-\tau} \operatorname{asinh}(Z) + \lambda). \quad (\text{S.3})$$

817 Thus, the two parameterizations are related by

818
$$\delta = e^\tau, \epsilon = \delta\lambda = e^\tau\lambda. \quad (\text{S.4})$$

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

826 S.2 Estimating mixed-effects models using shrinkage

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

834 Specialized software is often used to fit such models, such as the **nlme**, **lme4**, **mgcv**
835 and **gamm4** libraries in R, but these only allow a small subset of the distribution families
836 we want to consider for modeling growth increments (the **gamLss** package allows many
837 distribution families, but in our experience, even when random effects are simple in
838 structure the fitting algorithms often fail to converge or fail to find the global optimum).

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

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

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

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

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

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

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

872 `LogLik=function(pars,new.size,U){`

```

873 pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];
874 mu = U%*%pars1;
875 sigma = exp(pars2[1]+pars2[2]*mu);
876 dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)
877 }

```

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

```

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

```

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

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

$$890 \quad \hat{a}_t = a_t + \varepsilon_t \quad (\text{S.5})$$

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

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

$$901 \quad \mathbb{E}(\varepsilon_t | a_t) = 0. \quad (\text{S.6})$$

902 These are optimistic assumptions, but not excessively optimistic. Some degree of tem-
903 poral correlation will often be present, and as we explain at the end, it is theoretically
904 possible to account for it. Maximum likelihood parameter estimates are not unbiased,
905 but if the assumptions of maximum likelihood are satisfied the bias is asymptotically

906 negligible compared to the standard error (the bias scales as the inverse of sample size,
 907 the standard error as the square root of the inverse of sample size).

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

$$909 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S.7})$$

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

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

$$917 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S.8})$$

918 $\hat{\sigma}^2$ estimates the variance of the distribution from which the a_t are assumed to be drawn.

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

$$923 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.9})$$

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

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

$$931 \quad \tilde{a}_t = \bar{a}_t + \frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}} (\hat{a}_t - \bar{a}_t). \quad (\text{S.10})$$

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

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

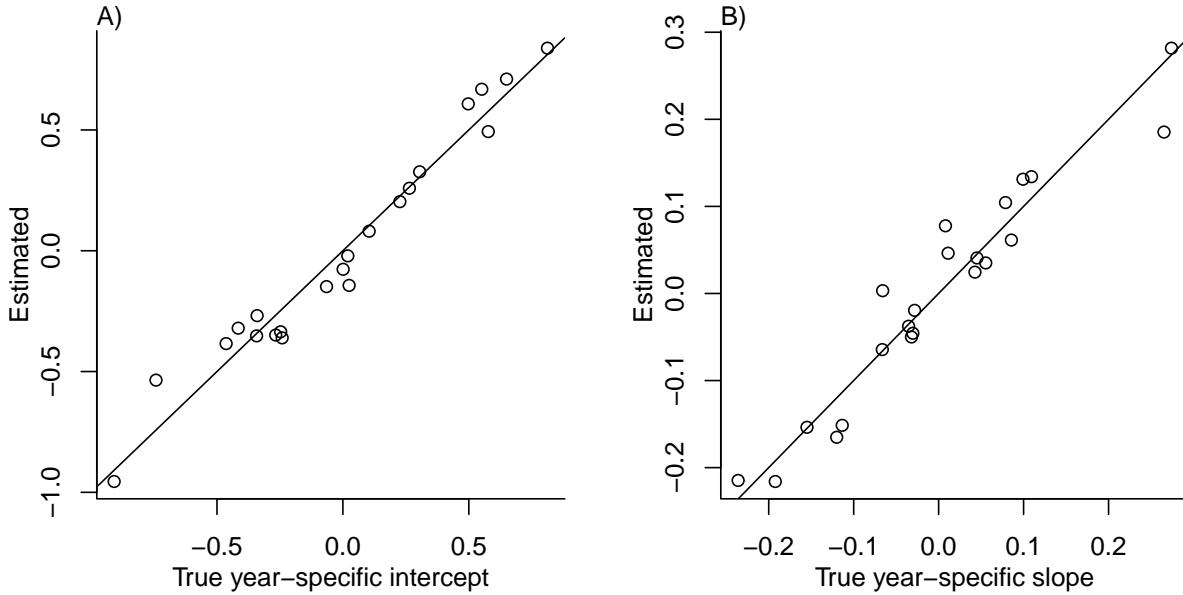
942 Finally, as mentioned above, the estimate of σ^2 can account for temporal autocor-
 943 relation in the a_t . When present, those correlations add a term to eqn. (S.7) (see eqn.
 944 (1) in Gould and Nichols (1998)), which can be estimated from the sample autocorre-
 945 lation of the \hat{a}_t .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

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

```
952 # Variance-covariance matrices for intercepts and slopes
953 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
954 # Extract year-specific intercepts, center them to zero
955 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
956
957 # Estimate sigma^2
958 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
959             (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
960
961 # Shrink deviations from the mean
962 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
963
964 # Do it all again for the slopes
965 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
966 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
967             (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
```

```
968 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));
```

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



975 **S.3 Additional case studies**

976 **S.3.1 Case study: Sea fan corals, *Gorgonia ventalina***

977 Bruno et al. (2011) developed an IPM to understand the rise and fall of a fungal pathogen
978 *Aspergillus sydowii* in Caribbean sea fan corals *G. ventalina*. The model was based on re-
979 peated observations of marked corals in permanent transects at several sites near Aku-
980 mal, Mexico, recording disease status (infected/uninfected) and the area of uninfected
981 tissue. The epidemic peak had passed and disease incidence was already low, so in-
982 fected fans were relatively infrequent. We therefore limit the analysis here to uninfected
983 individuals. Bruno et al. (2011) found statistically significant year and site effects, but
984 as those explained a very small fraction of the variation in growth increments, they
985 fitted a single growth model to data pooled across years and sites. We do the same
986 here. The pooled data set consists of 358 observed size transitions. The data exhibited

987 size-dependent variance in growth (change in area, cm^2). Bruno et al. (2011) chose to sta-
988 bilize the variance by cube-root transforming size, and then fitting the standard model
989 with Gaussian growth increments. Here we take a different approach, using natural log
990 transformation of area and modeling size-dependent variance.

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

```
996 # XH is a data frame holding the data
997 # logarea.t0, .t1 denote initial and final values of log-transformed area
998 fitGAU <- gam(list(logarea.t1~ s(logarea.t0), ~ s(logarea.t0)),
999   data=XH, gamma=1.4, family=gaulss())
1000 fitted_all = predict(fitGAU,type="response");
1001 fitted_sd = 1/fitted_all[,2];
1002 scaledResids = residuals(fitGAU,type='response')/fitted_sd;
```

1003 Fig. S-1A shows the log-transformed data and Gaussian model. The mean function
1004 (solid red curve) is visually nearly linear, but the fitted spline is strongly favored over a
1005 linear model for the mean ($\Delta AIC \approx 9$). The spline for standard deviation σ versus initial
1006 size reflects the evident greater variability in growth at smaller sizes.

1007 There are no blatant signs of trouble in the pilot Gaussian model, but quantile re-
1008 gressions on the scaled residuals, and the NP Skewness and Kurtosis metrics derived
1009 from them (Eq. 3 and 4), suggest deviations from normality (Fig. S-1B). Specifically,
1010 skewness switches from negative to positive across the size range, with smaller corals
1011 more prone to extreme shrinkage and larger corals more prone to extreme growth. Kur-
1012 tosis also changes direction over the size distribution, with thinner tails than Gaussian
1013 at small sizes and fatter tails at large sizes. The fitted nonparametric moments suggest
1014 that the upper and lower tails of size transition probabilities may differ by up to 20%,
1015 and the weight of the tails may be 20% greater or less than Gaussian, depending on
1016 initial size – not overwhelming deficiencies, but not trivial either. Are these deviations
1017 from normality severe enough to warrant a second, non-Gaussian iteration of growth
1018 modeling? To answer that question, we simulated data from the fitted Gaussian model
1019 and examined whether key properties of the simulated data are consistent with those
1020 of the real data – this is the ultimate litmus test for a growth model's adequacy and
1021 should be a standard element of IPM construction, in our opinion. If the simulated data

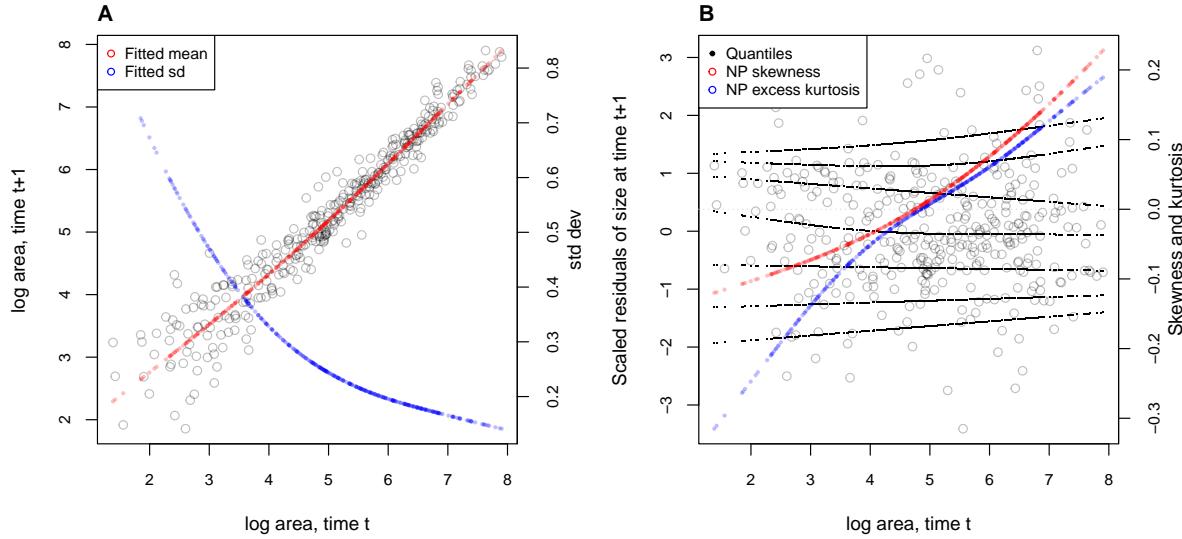


Figure S-1: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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.

1022 are not consistent with the real data, it is time to choose a better distribution (Fig. 1).
 1023 In this case, most of 100 Gaussian model simulations are out of line with the skew at
 1024 smallest and largest sizes, and excess kurtosis observed at moderately large sizes (Fig.
 1025 S-2 CD). For at least some parts of the size distribution, a non-Gaussian model would
 1026 better capture size transitions.

1027 We sought a distribution that could accommodate the observed changes in the sign
 1028 of skewness and excess kurtosis. We chose the sinh-arcsinh (SHASH) distribution, a
 1029 four-parameter distribution that, conveniently, is included in **mgcv**'s **gam()** function.
 1030 For consistency with the Gaussian for location and scale, specification of basis functions
 1031 ($k = 4$) is limited to parameters for skewness and kurtosis:

```
1032 fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0), # <- location
1033 ~ s(logarea.t0), # <- log-scale
1034 ~ s(logarea.t0,k=4), # <- skewness
1035 ~ s(logarea.t0,k=4)), # <- log-kurtosis
1036 data = XH, gamma = 1.4, family = shash, optimizer = "efs")
```

1037 The fitted model's mean and variance are nearly identical to the Gaussian (Fig. S-2AB),
 1038 and the fitted trends in skewness and kurtosis are much less "wiggly" than the estimate

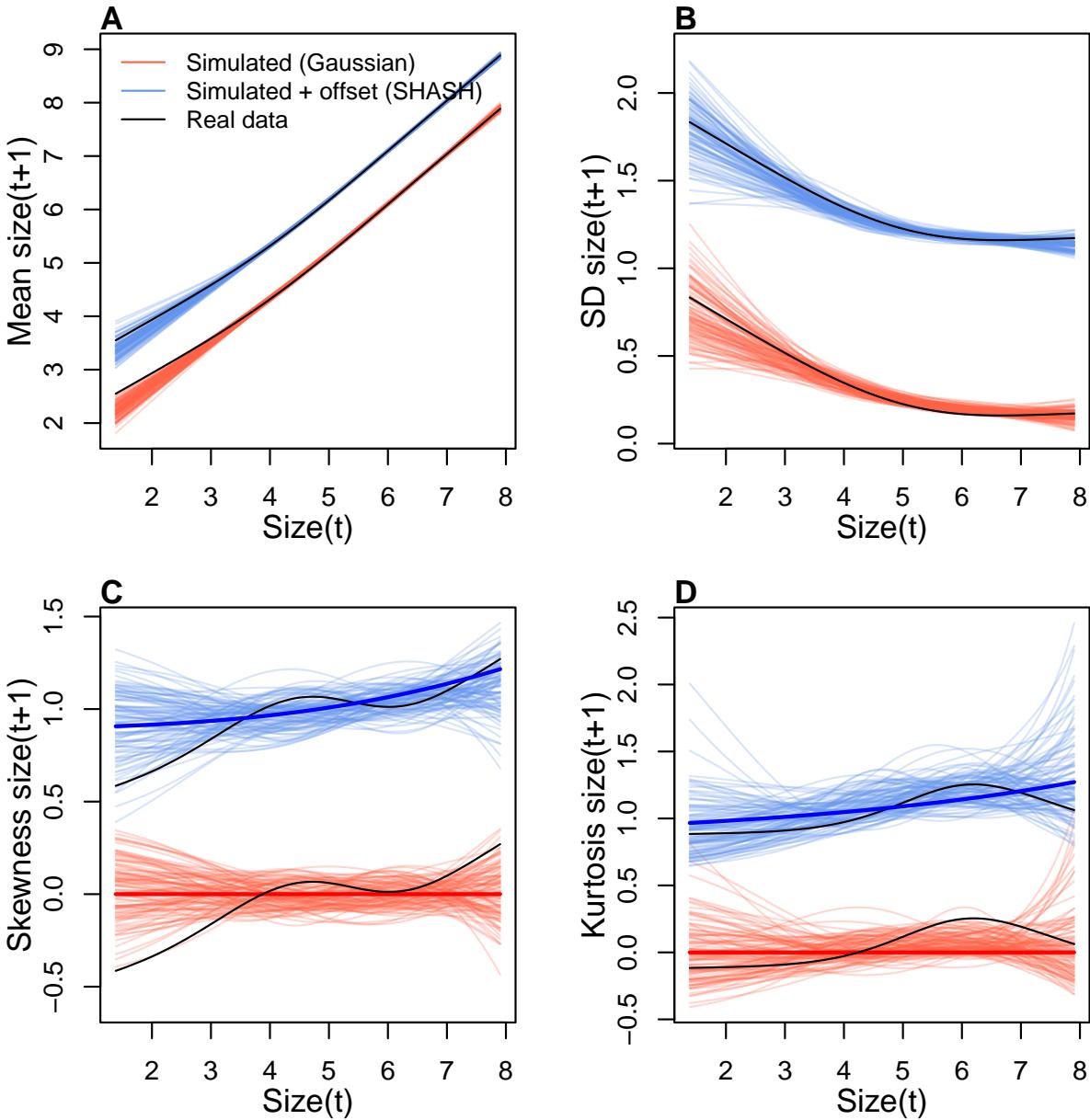


Figure S-2: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Note that plotted values for the SHASH are offset by one unit to allow comparisons. In the skewness and kurtosis panels, the darker solid curves show the values for the fitted growth models. Figure made by script AkumalCorals_qgam.R.

from the data (Fig. S-2CD). Nonetheless, data simulated from the SHASH model are more consistent with the real data, with more SHASH data sets matching or exceeding the largest skewness and kurtosis values observed (Fig. S-2CD). If one cares to quantify

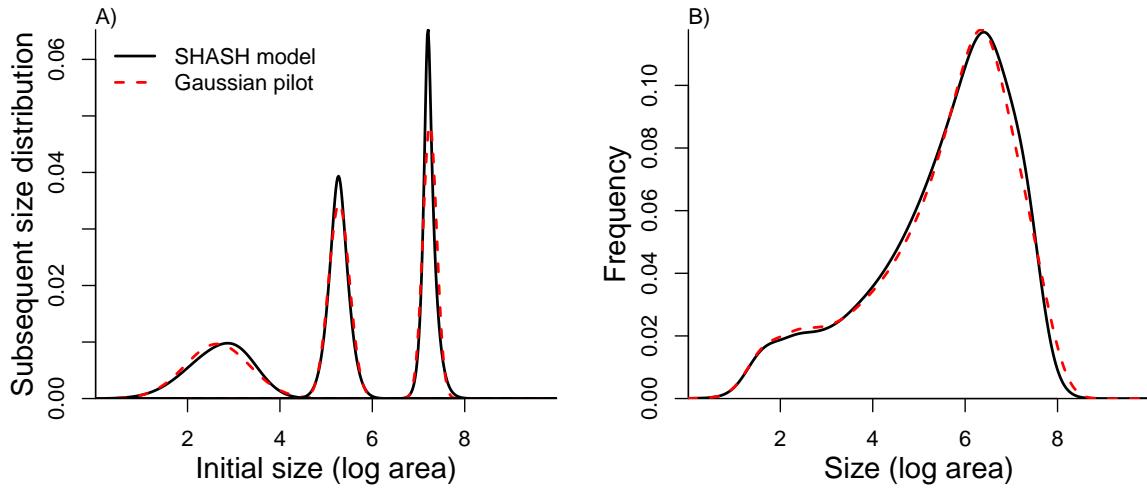


Figure S-3: Comparisons between the fitted SHASH growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year $t + 1$ for three different values of size in year t . The leftmost pair of curves are for initial size equal to the median size of a new recruit (data from (Bruno et al., 2011)); central pair is for the median initial size of uninfected individuals; rightmost pair are for the 95th percentile initial size for uninfected individuals. B) Steady-state size distributions resulting from a constant unit input of new recruits. As in Bruno et al. (2011) we assume that larvae are very widely dispersed, so that the number of recruits arriving at any one location is independent of the local population abundance or structure. Size distribution of recruits was described by a kernel density estimate based on the measured sizes of known new recruits ($n = 9$). Figure made by script AkumalCoralsIPMs.R.

the difference between models, the SHASH model is clearly favored by AIC ($\Delta AIC = 5.45$) despite having twice as many parameters to fit.

What, then, have we gained by fitting a better growth model? Fig. S-3A compares the predicted distributions of subsequent size in the fitted model and Gaussian pilot models, for the median size of a new recruit (leftmost pair of curves), the median initial size (central curves), and the 95th percentile of initial size in the data (rightmost curves). The differences are small, and most pronounced for the smallest size, where recruits are predicted to grow slightly larger under the SHASH model than the Gaussian model. The direction of this difference was surprising, because the SHASH has negative skew at small sizes in the data. However, the SHASH model also gives a better prediction of mean growth at small sizes than the Gaussian model. At intermediate sizes the predictions are nearly identical; at large sizes the SHASH has slightly lower standard deviation, but fatter tails (excess kurtosis). Fig. S-3B shows the predicted steady-state size distributions resulting from a constant unit input of recruits. Again, the differences

1056 are very subtle. Finally, the Gaussian and SHASH growth models predict very similar
1057 mean life span (17.7 and 17.9 years, respectively).

1058 From these outputs, there is little evidence that improved modeling of coral growth
1059 meaningfully improved biological inferences from the IPM. One could argue that it was
1060 not worth the trouble, even though it was almost no trouble at all. But before fitting
1061 the SHASH model, we could not have known whether or not it would have made a
1062 difference.

1063 In this case study we used `gam` to fit both the Gaussian and SHASH models because
1064 that obviated model selection on functions for mean, variance, and higher moments.
1065 However, `gam` should be used with caution. Nonparametric regression models notori-
1066 ously “wag their tails” because the ends of the fitted curve can be pulled close to the
1067 outermost data points. This is especially problematic for growth modeling, because data
1068 are typically sparse near the bounds of the size distribution. To minimize the risk of
1069 overfitting we specified the number of “knots” ($k=4$) and used $\text{gamma}=1.4$ to overweight
1070 model degrees of freedom as suggested by Gu (2013, sec. 3.2). But it is always impor-
1071 tant to plot the fitted splines and make sure they do not wag unrealistically. If they do,
1072 parametric regression may be a better choice.

1073 S.3.2 Case study: creosotebush, *Larrea tridentata*

1074 Our next case study comes from our studies of the woody shrub creosotebush (*Larrea tri-*
1075 *dentata*) at the Sevilleta Long-Term Ecological Research (LTER) site in central New Mex-
1076 ico, US. At this site as elsewhere in the Southwest US, creosotebush is encroaching into
1077 desert grassland habitats. The data described here were collected along transects span-
1078 ning grass-shrub ecotones to understand patterns of density dependence in creosotebush
1079 demography. Specifically, we asked whether fitness is maximized approaching zero den-
1080 sity at the leading edge of the expansion front (consistent with ‘pulled’ expansion), or
1081 whether there is a demographic advantage for shrubs at higher density due to positive
1082 feedbacks expected for ecosystem engineers (leading to ‘pushed’ expansion). Our pub-
1083 lished study (Drees et al., 2023) used a spatial integral projection model (SIPM) to predict
1084 the speed of shrub encroachment, assuming normally-distributed size transitions. Here
1085 we step through our suggested workflow to ask whether a non-Gaussian model would
1086 have been more faithful to the data, and how such an improvement would influence
1087 predictions for the speed of encroachment.

1088 Growth data come from 522 shrubs censused longitudinally over four years (2013-
1089 2017). Census individuals occurred along 12 replicate transects (200 to 600 m in length)

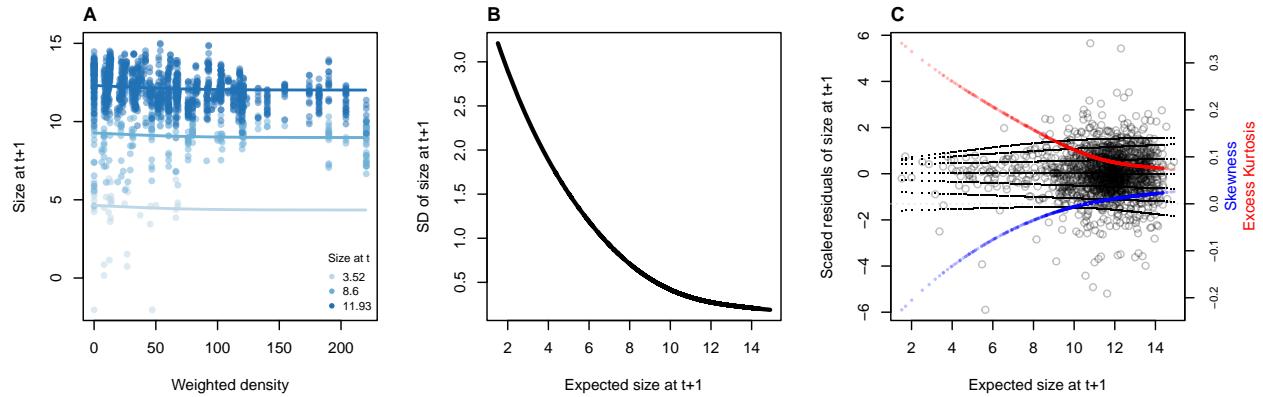


Figure S-4: **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume (cm^3). **B**, Standard deviation of size at time $t + 1$ 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. All figures made by script `creosote_growth_modeling.R`.

that spanned gradients of shrub density along shrub-grass ecotones. Size was measured as volume of an elliptical cone based on height and width measurements; the size variable of the IPM was the natural logarithm of volume (cm^3). For each census individual, we recorded the size and density of all conspecifics within the five-meter transect “window” in which it occurred, and took the sum of all sizes within the window as a weighted measure of local density. The data are available in Ochocki et al. (2023).

As an initial Gaussian approach, and following the approach of Drees et al. 2023, we first fit a generalized additive model with `mgcv` that included smooth terms for initial size and weighted density (constrained to four basis functions), plus the random effect of transect. We used the `gaulss` family and, as a starting point, fit a constant standard deviation.

```
1101 LATR_GAU <- gam(list(log_volume_t1~ s(log_volume_t,k=4) +
1102   s(dens_scaled,k=4) + s(unique.transect,bs="re"), ~ 1),
1103   family="gaulss", data=LATR_grow, method="ML", gamma=1.4)
```

Using the fitted values from this initial model, we updated the standard deviation to be a smooth function of fitted values, and iterated the fitting until the weights stopped changing, following the same steps as in the orchid case study.

The resulting Gaussian growth model predicts strong initial size-dependence and weak and slightly nonlinear (but monotonic) negative density dependence (Fig. S-4A).

1109 The model accounts for non-constant variance, which indicate greater dispersion for
1110 smaller values of expected size (Fig. S-4B). Quantiles of the standardized residuals indi-
1111 cate that skew and excess kurtosis are both greater at smaller sizes (Fig. S-4C). Skewness
1112 is close to zero for larger plants (the best-sampled size range) but excess kurtosis re-
1113 mains positive for large plants (ca. 10% heavier tails than Gaussian). As a candidate
1114 for improvement, we turned to the Johnson's S_U (JSU) distribution, a four-parameter,
1115 leptokurtic distribution capable of skew in either direction.

1116 Following our suggested workflow, rather than re-fitting a JSU model from scratch,
1117 we parameterize a model a model where the residuals from the Gaussian model are fitted
1118 by a JSU distribution. This is relatively easy because the **gamlss.dist** package provides
1119 a parameterization of the JSU in which the location parameter μ is the mean and scale
1120 parameter σ is the standard deviation (Rigby et al., 2019). We fit the "hybrid" model by
1121 writing a likelihood function that uses the fitted mean and standard deviation functions
1122 from Gaussian pilot model, and estimates the parameters that control skewness and
1123 kurtosis as linear functions of predicted future size. The "hybrid" likelihood looks like
1124 this:

```
1125 JSULogLik=function(pars){  
1126   dJSU(LATR_grow$log_volume_t1,  
1127     mu=LATR_grow$GAU_mean,  
1128     sigma=LATR_grow$GAU_sd,  
1129     nu = pars[1]+pars[2]*LATR_grow$GAU_mean,  
1130     tau = exp(pars[3]+pars[4]*LATR_grow$GAU_mean), log=TRUE)  
1131 }
```

1132 The mean and standard deviation of the JSU are set to those of the best Gaussian
1133 model. Based on diagnostics of the standardized residuals (Fig. S-4C), JSU parameters
1134 that control skewness and kurtosis are defined as linear functions of the mean, and those
1135 coefficients are estimated by maximum likelihood. This approach relies on the robust-
1136 ness of Gaussian models fitted mean and variance to deviations from normality. If one is
1137 skeptical of this approach, it is possible to simultaneously re-fit all parameters of the JSU
1138 in a maximum likelihood framework. However, incorporating random effects into a cus-
1139 tom likelihood model is non-trivial (we provide guidance on one way to do this, using
1140 the "shrinkage" approach, in Appendix XX). Therefore a key advantage of the hybrid
1141 approach is convenient retention of the fitted random effects and associated variance
1142 components, which get shuttled from the Gaussian model into the non-Gaussian model
1143 without any fuss (it was critical that we used a parameterization of the JSU for which

1144 mu is the mean and sigma is the standard deviation). And, if this approach does not
1145 “work” (i.e., deviations from normality biased the fitted values of the Gaussian model)
1146 one would quickly find out through the simulation step of the workflow.⁷ The hybrid
1147 JSU model performed well, generating simulated data that aligned with the real data
1148 better than the best Gaussian model, particularly in the standard deviation and kurtosis
1149 (Fig. S-5). The JSU model has exactly the same mean and standard deviation of fu-
1150 ture size as the Gaussian, but Fig. S-4 uses the quantile-based nonparametric mean and
1151 standard deviation. The results show that even though the JSU was not fitted to match
1152 those, it comes closer than the Gaussian model as a result of accounting for the skew
1153 and kurtosis.

1154 The improvement of the JSU over the Gaussian growth model, while visually sat-
1155 isfying, had only weak influence on SIPM results. The Gaussian model slightly over-
1156 estimated the low-density growth rate, but models using either Gaussian or JSU growth
1157 kernels had very similar monotonic decreases in λ with increasing local density, and
1158 nearly identical wave velocities (Fig. S-6). This species has very low mortality risk once
1159 established (mean remaining life expectancy of a median-sized shrub is 24,408 years)
1160 and its population growth and wave expansion are limited by very low seedling recruit-
1161 ment ((Drees et al., 2023)). Weak size-dependence in survival likely explains why the
1162 improvement in growth modeling had little influence on SIPM predictions.

7Move to orchids.

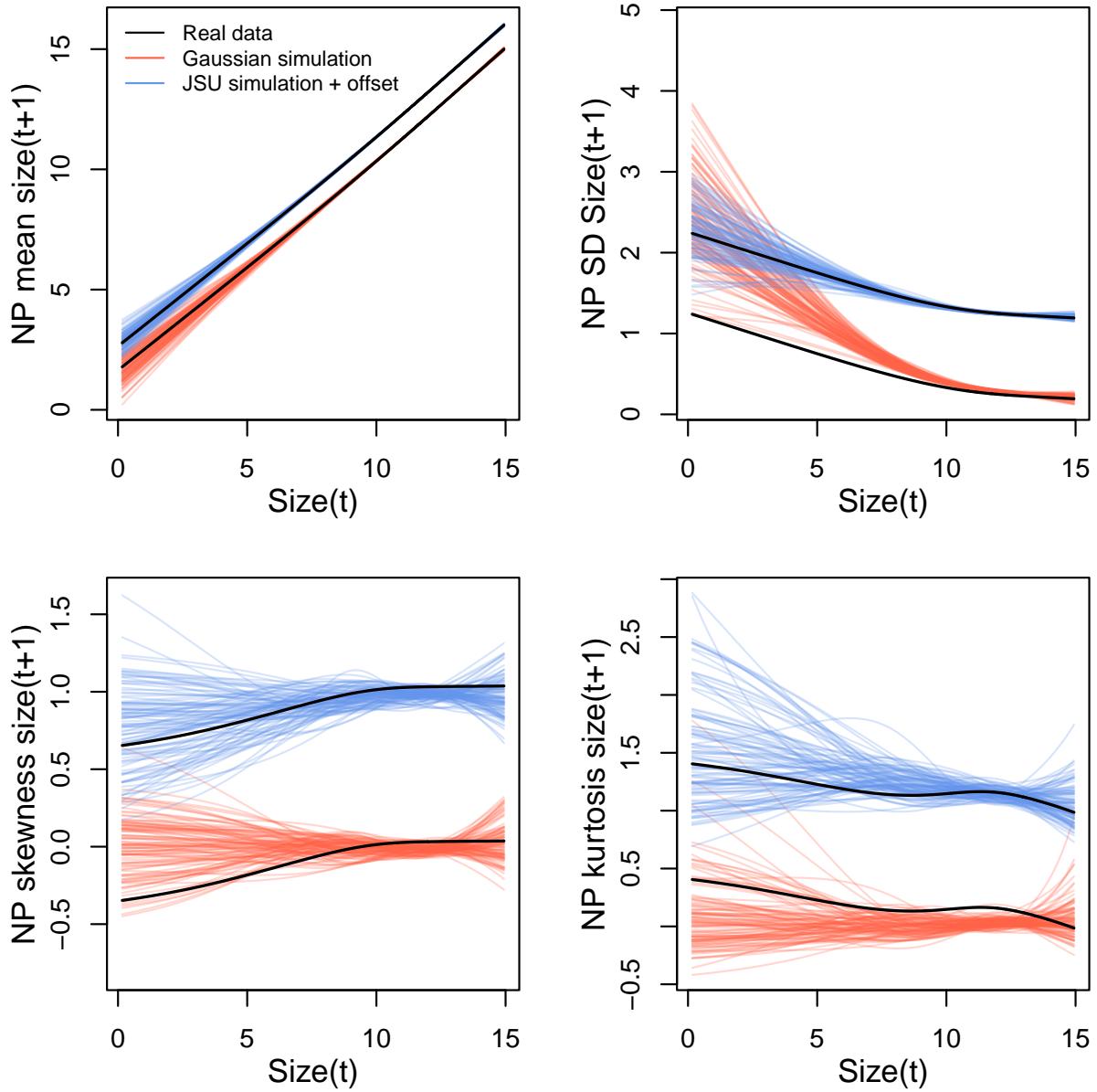


Figure S-5: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect to initial size; their distribution is also conditional on density but initial size is by far the stronger predictor of future size, so we chose this visualization. Values for the JSU model (and the corresponding “real data” values) are offset vertically by one unit for comparison. Figure made by script `creosote_growth_modeling.R`.

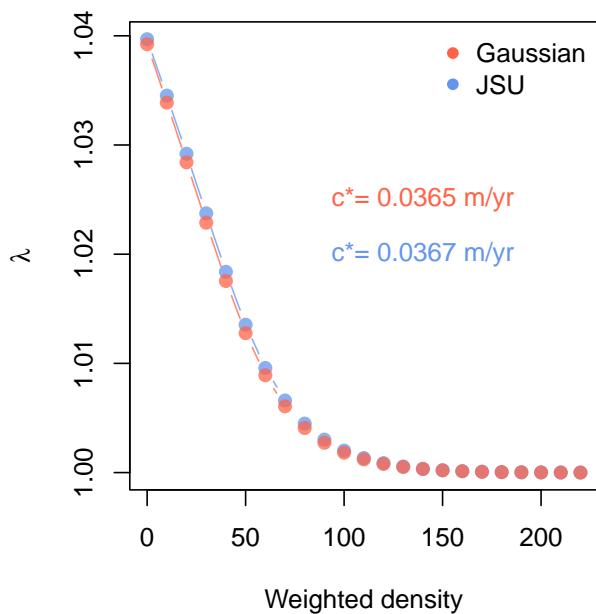


Figure S-6: Density dependence in fitness (λ) and asymptotic velocity of the creosote encroachment wave (c^*) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ($\log(cm^3)$) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.