

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

Tom E.X. Miller<sup>\*a</sup> and Stephen P. Ellner<sup>b</sup>

<sup>a</sup>Department of BioSciences, Rice University, Houston TX

<sup>b</sup>Department of Ecology & Evolutionary Biology, Cornell University, Ithaca NY

**Submitted to:** *Ecology* (Statistical Report)

**Keywords:** demography; growth; integral projection model; kurtosis; skewness

**Open Research Statement:** Data are already published and publicly available, with those items properly cited in this submission. Three data sets are cited as data packages (Ochocki *et al.*, 2023; Winfield *et al.*, 2013b; Miller, 2020). Two other data sets are available in our Github repo, which also includes all of our code. The repo will be archived in a Zenodo package, with the DOI included at the end of the article, upon publication. During peer review, our data and code are available at [https://github.com/texmiller/IPM\\_size\\_transitions](https://github.com/texmiller/IPM_size_transitions).

---

\*Corresponding author. Department of BioSciences, Rice University, Houston, TX 77005-1827. Email: tom.miller@rice.edu  
Phone: 713-348-4218

## **1 Abstract**

2 Integral Projection Models (IPMs) are widely used for studying continuously size-structured populations.

3 IPMs require a growth sub-model that describes the probability of future size conditional on current size and

4 any covariates. Most IPM studies assume that this distribution is Gaussian, despite calls for non-Gaussian

5 models that accommodate skewness and excess kurtosis. We provide a general workflow for accommodating

6 non-Gaussian growth patterns while retaining important covariates and random effects. Our approach

7 emphasizes visual diagnostics from pilot Gaussian models and quantile-based metrics of skewness and kurtosis

8 that guide selection of a non-Gaussian alternative, if necessary. Across five case studies, skewness and excess

9 kurtosis were common features of growth data and non-Gaussian models consistently generated simulated data

10 that were more consistent with real data than pilot Gaussian models. However, effects of “improved” growth

11 modeling on IPM results were moderate to weak, and differed in direction or magnitude between different

12 outputs from the same model. Using tools not available when IPMs were first developed, it is now possible to fit

13 non-Gaussian models to growth data without sacrificing ecological complexity. Doing so, as guided by careful

14 interrogation of the data, will result in models that better represent the populations for which they are intended.

15 **1 Introduction**

16 Structured demographic models – matrix and integral projection models (MPMs and IPMs) – are powerful  
17 tools for data-driven modeling of population and community dynamics. In contrast to MPMs for populations  
18 with discrete structure (life stage, age class, etc.), IPMs (Easterling *et al.*, 2000) accommodate populations  
19 structured by continuous state variables, most commonly size. A related innovation of the IPM framework  
20 is its emphasis on regression-based modeling for parameter estimation, which often carries important  
21 advantages for making the most of hard-won data (Ellner *et al.*, 2022).

22 A standard workflow allows ecologists to assemble an IPM from data using familiar regression  
23 tools to describe growth, survival, reproduction, and other demographic transitions as functions of size  
24 (Coulson, 2012; Ellner *et al.*, 2016). The relative ease of regression analyses, accommodating covariates  
25 (e.g., environmental factors, experimental treatments) and complex variance structures (e.g., random effects,  
26 correlated errors), has facilitated a growing IPM literature that examines how biotic or abiotic factors affect  
27 population dynamics (e.g., Louthan *et al.*, 2022; Ozgul *et al.*, 2010) and explores the consequences of  
28 demographic heterogeneity associated with spatial, temporal, and individual variation (e.g., Compagnoni *et al.*,  
29 2016; Crone, 2016; Plard *et al.*, 2018). The vital rate regressions (or “sub-models”) are the bridge between  
30 the individual-level data and the population-level model and its predictions; it is important to get those right.

31 Compared to other vital rates, growth is special. The survival and reproduction sub-models only  
32 need to provide a single predicted value as functions of size (we use “size” as the name for whatever  
33 continuous variable defines the population structure). But the growth model must specify the full probability  
34 distribution of subsequent size conditional on initial size, defining the growth ‘kernel’  $G(z', z)$  that gives the  
35 probability density of future size  $z'$  at time  $t+1$  conditional on current size  $z$  at time  $t$ . Whenever survival  
36 and reproduction are size-dependent, the entire distribution of size transitions can strongly influence IPM  
37 predictions because it governs how frequently size changes are much greater or much lower than average.

38 Easterling et al. 2000 provided the original template for modeling size transitions in IPMs. They  
39 first tried simple linear regression, assuming Normally distributed size changes with constant variance.  
40 Because the residuals from this regression exhibited non-constant variance, they used a two-step approach  
41 to estimate the size-dependence in mean squared residuals (better options soon became available, such  
42 as the `lme` function in R). However, even after accounting for non-constant variance, growth data may  
43 still be non-Normal. Size transitions are often skewed such that large decreases are more common than  
44 large increases (Peterson *et al.*, 2019; Salguero-Gómez & Casper, 2010), or vice versa (Stubberud *et al.*,  
45 2019). Size transitions may also exhibit excess kurtosis (“fat tails”), where extreme growth or shrinkage  
46 is more common than predicted by the tails of the Normal distribution (Hérault *et al.*, 2011).

47 The observation that the Normal (or Gaussian) distribution may poorly describe size transitions in real  
48 organisms has been made before, and several studies have emphasized that alternative distributions should be  
49 explored (Easterling *et al.*, 2000; Peterson *et al.*, 2019; Rees *et al.*, 2014; Williams *et al.*, 2012). For example,  
50 Peterson et al. 2019 showed that skewness in size transitions could be modeled through beta regression  
51 on transformed data (for reasons we describe below, this approach also has some drawbacks), or by fitting  
52 a skewed Normal distribution. They showed that incorporating skew could have important consequences for  
53 model-based inferences, and concluded that “testing of alternative distributions for growth... [should] become  
54 standard in the construction of size-structured population models.” Nonetheless, default use of Gaussian  
55 growth distributions (often with non-constant variance) remains the standard practice. The general state-of-  
56 the-art in the literature appears to remain where it was 20 or so years ago, using the default Gaussian model  
57 without examining critically whether or not it actually describes the data well. We are guilty of this, ourselves.

58 The persistence of Gaussian growth models is understandable. Popular packages such as `lme4` (Bates  
59 *et al.*, 2015), `mgcv` (Wood, 2017), and `MCMCglmm` (Hadfield *et al.*, 2010) make it easy to fit growth  
60 models with potentially complex fixed- and random-effect structures, but the possible distributions of  
61 continuous responses are limited, and default to Gaussian. Abandoning these convenient tools for the

62 sake of more flexible growth modeling means, it may seem, sacrificing the flexibility to model diverse  
63 sources of demographic variation, some of which may be the motivation driving the study in the first place.

64 Our goal here is to present and illustrate a practical “recipe” that moves growth modeling past the  
65 standards set over 20 years ago. Using software tools that are now readily accessible, ecologists can escape  
66 the apparent trade-off between realistically modeling non-Gaussian size transitions and flexibly including  
67 multiple covariates and random effects.<sup>1</sup> As with any recipe, users may need to make substitutions or  
68 add ingredients to suit their needs. We emphasize graphical diagnostics for developing and evaluating  
69 growth models, rather than a process centered on statistical tests or model selection. Through empirical  
70 case studies we demonstrate how tools that were nonexistent or not readily available when IPMs first came  
71 into use now make it straightforward and relatively easy to identify when the default model is a poor fit  
72 to the data, and to then choose and fit a better growth model that is no harder to use in practice. We illustrate  
73 our approach by revisiting three published case studies (and three additional case studies in the Supporting  
74 Information), including examples from our own previous work. In each case, the Gaussian assumption  
75 does not stand up to close scrutiny. We illustrate how we could have done better, and the consequences of  
76 “doing better” for our ecological inferences. All analyses were carried out in R (R Core Team, 2022) version  
77 4.0 or higher and may be reproduced from publicly available code and data (see *Data Availability Statement*).

## 78 **2 Flexible growth modeling**

79 The modeling process that we suggest runs as follows (Fig. 1):

80 **1. Fit a “pilot” model assuming a Gaussian distribution, but allowing for non-constant variance.** This  
81 step is familiar to most IPM users, as it is the start and end of the standard approach. It may include model  
82 selection to identify which treatment effects or environmental drivers affect the mean and/or variance of future  
83 size. Non-constant variance is often fitted in a two-stage process, first fitting mean growth assuming constant

---

<sup>1</sup>Our statements about software availability are based on what current software reliably delivers in our personal experience, not on what they promise.

84 variance, then doing a regression relating the squared residuals to initial size or the fitted mean of subsequent  
85 size. Fitting mean and variance simultaneously as functions of initial size, as can be done with R packages  
86 **mgev** and **nmle**, is advantageous when possible because incorrectly assuming constant variance can affect  
87 model selection for the mean. We illustrate both one-step and two-step approaches in the case studies below.

88 Allowing non-constant variance removes the need for transforming the data to stabilize growth variance.  
89 Transformation may still be useful if it does not create new problems such as making some state-fate  
90 relationships highly nonlinear. In particular, log-transformation often reduces or eliminates heteroskedasticity  
91 in growth data (Ellner *et al.*, 2016) and also helps avoid eviction at small sizes (Williams *et al.*, 2012).

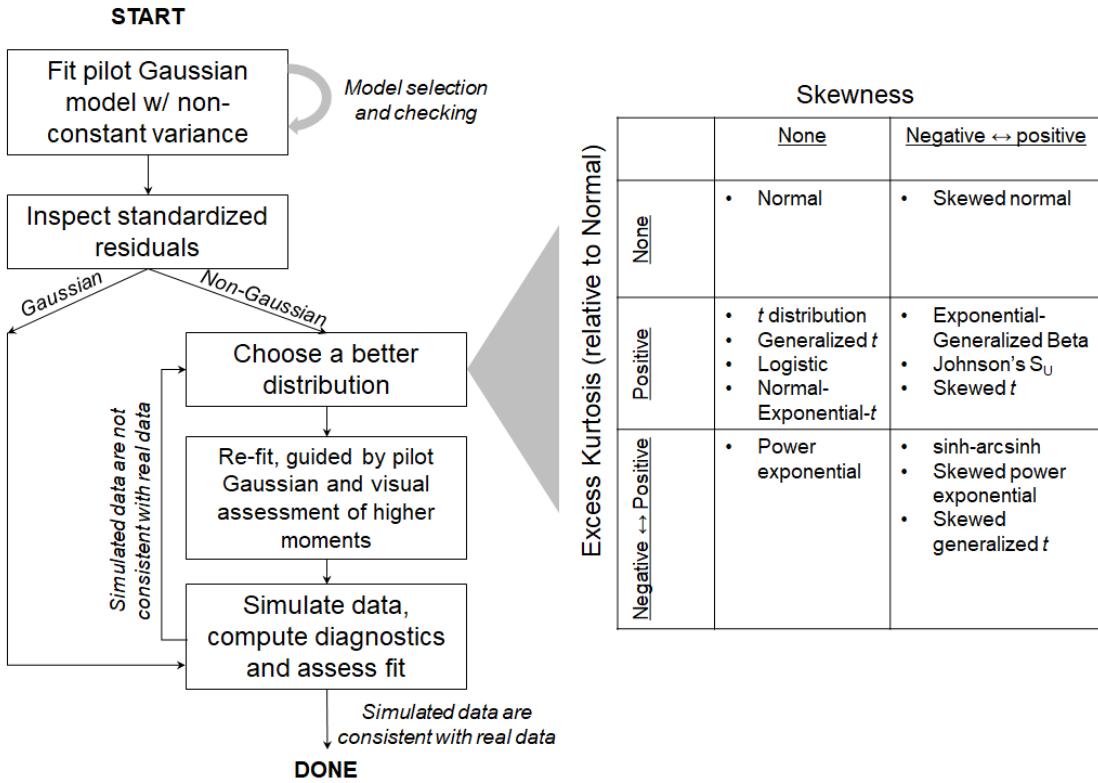
92 The fitted mean and variance functions should be checked before going any further. If they are  
93 perfectly correct, standardized residuals (residuals scaled by the standard deviation) will have zero mean  
94 and unit variance overall, and will exhibit no trends in mean or variance with initial size or fitted mean  
95 value. However, estimates of the mean and variance functions are somewhat smoothed because of the  
96 inescapable bias-variance tradeoff, so scaled residuals will retain some variation in location and scale.  
97 Given enough data, statistical tests will detect that variation. So instead, we take for granted the presence  
98 of trends and assess their importance by fitting nonparametric spline regression models for residuals (trend  
99 in mean) and absolute residuals (trend in variance) as a function of initial size or fitted value. The mean  
100 and variance functions can be accepted if the regression curves for the scaled residuals are nearly flat.

101 **2. Use graphical diagnostics to identify if and how the standardized residuals deviate from Gaussian,**  
102 **and to choose a more appropriate distribution.** If the Gaussian growth model is valid, the standardized  
103 residuals should be Gaussian with zero skewness or excess kurtosis. Growth data may deviate from this in many  
104 ways, and the nature of the deviations can guide the search for a better distribution. Tests such as the D'Agostino  
105 test of skewness (D'Agostino, 1970) and the Anscombe-Glynn test of kurtosis (Anscombe & Glynn, 1983)  
106 can be used to diagnose whether the standardized residuals, in aggregate, deviate from normality (Komsta &

<sup>107</sup> Novomestky, 2015). However, the aggregate distribution may be misleading if skewness or kurtosis vary with  
<sup>108</sup> size or other covariates. Skewness changing from positive at small sizes to negative at large sizes might produce  
<sup>109</sup> zero overall skewness, but really requires a distribution that can allow both positive and negative skew, such  
<sup>110</sup> as the skewed Normal or Johnson  $S_U$  distributions. Alternatively, growth data may exhibit leptokurtosis (in  
<sup>111</sup> which case the  $t$  distribution may be a good choice) or may shift from platykurtosis to leptokurtosis depending  
<sup>112</sup> on initial size (in which case the power exponential distribution may be a good choice). It is therefore essential  
<sup>113</sup> to visualize trends in distribution properties with respect to either initial size, or expected future size for models  
<sup>114</sup> with multiple covariates. Fig. 1 includes guidance on how the skew and kurtosis properties of the standardized  
<sup>115</sup> residuals suggest options for an appropriate growth distribution. In our case studies we exploit the many  
<sup>116</sup> distributions in the **gamlss** R package (Stasinopoulos *et al.*, 2007), but other distribution families can be used.

<sup>117</sup> **3. Refit the growth model using the chosen distribution.** In models with multiple covariates and/or random  
<sup>118</sup> effects, each potentially affecting several distribution parameters, “refit the model” could entail a massive  
<sup>119</sup> model selection process to identify the “best” non-Gaussian model. With so many options, model uncertainty  
<sup>120</sup> may be overwhelming and over-fitting becomes a significant risk even when precautions against it are taken.

<sup>121</sup> We therefore argue for adopting a more modest goal: remedy the defects evident in the standardized  
<sup>122</sup> residuals of the Gaussian model. This recommendation is based on the finding that parameter estimation  
<sup>123</sup> using Gaussian regression models is generally robust to deviations from normality of the residuals (Scheipl et al.,  
<sup>124</sup> 2020). That is, the fitted mean of the Gaussian model (as a function of covariates) is probably a  
<sup>125</sup> very good approximation for the fitted mean in the corresponding non-Gaussian model — and if it is  
<sup>126</sup> not, the next step in the modeling process will catch that. The functional forms for skew and kurtosis  
<sup>127</sup> of the non-Gaussian model can be guided by the qualitative features of the graphical diagnostics (e.g.,  
<sup>128</sup> that skewness switches from positive to negative with increasing size). As we demonstrate below, the  
<sup>129</sup> mean and standard deviation functions can often be carried over exactly from the pilot Gaussian model.



**Figure 1:** Recommended steps in 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 R package **gamlss.dist**, except for the skewed generalized  $t$ , which is available in the package **sgt** (Davis, 2015).

- 130 **4. Evaluate the final growth model through graphical diagnostics comparing simulated and real**
- 131 **growth data.** A good model will generate simulated data that look like the real data. Again, it is important
- 132 to inspect the properties of simulated data as a function of initial size, fitted mean, or other covariates
- 133 rather than examining the aggregate distribution. We again suggest below graphical diagnostics, based
- 134 mainly on quantiles, that can be used to compare simulated with real growth data. If the simulated data do
- 135 not correspond well with the real data, alternative or more flexible distribution families should be considered,
- 136 or more complex functions relating distribution parameters to size and other covariates.

<sup>137</sup> **3 How should skewness and kurtosis be measured?**

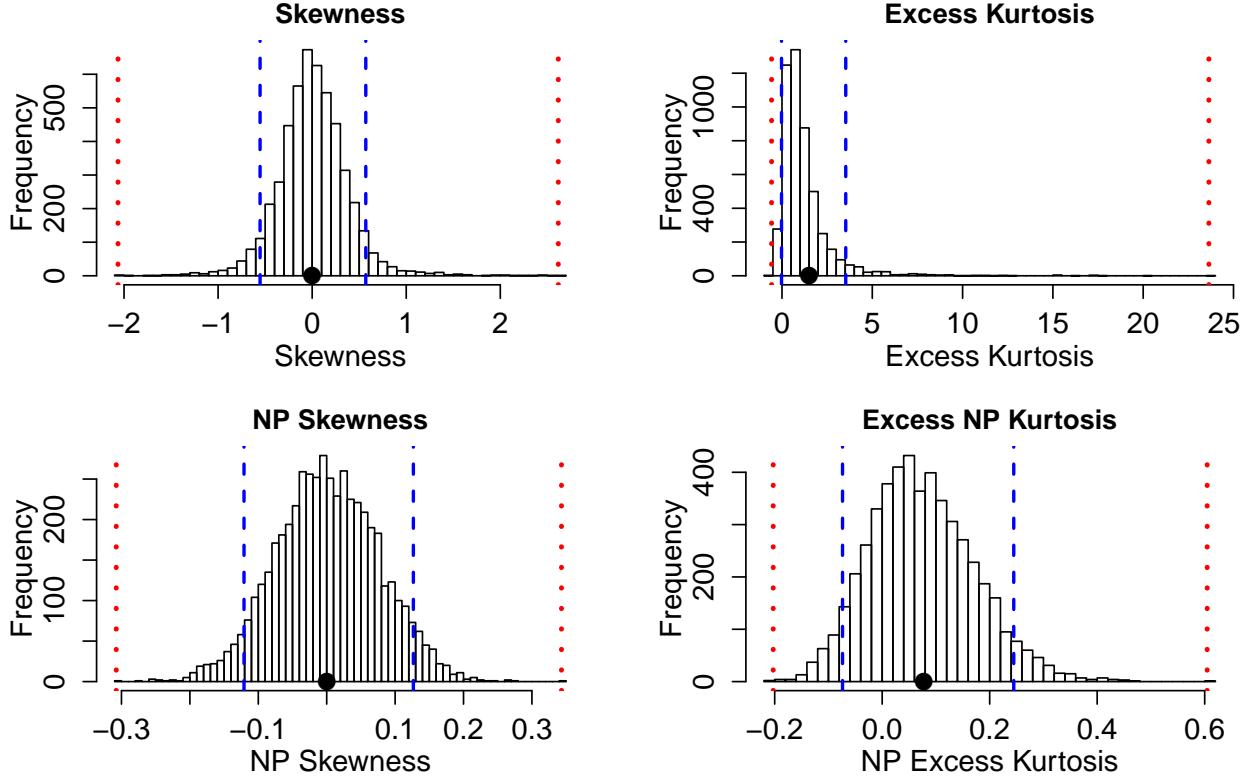
<sup>138</sup> Non-Gaussian growth modeling requires scrutinizing the skewness and kurtosis of standardized residuals,  
<sup>139</sup> so measurement of these properties warrants attention. The standard measures are based on the third and  
<sup>140</sup> fourth central moments, respectively, of the distribution: skewness =  $m_3/\sigma^3$ , excess kurtosis =  $m_4/\sigma^4 - 3$   
<sup>141</sup> where  $m_k = \mathbb{E}(X - \bar{X})^k$  is the  $k^{th}$  central moment of a random variable  $X$  and  $\sigma^2$  is the variance (second  
<sup>142</sup> central moment). A Gaussian distribution has zero skewness and zero excess kurtosis.

<sup>143</sup> The standard measures are simple and easy to use, but they have poor sampling properties. Because  
<sup>144</sup> the measures involve high powers of data values, a few outliers can produce very inaccurate estimates. Figure  
<sup>145</sup> 2 shows a simulated example, where the underlying data are samples of 200 values from a  $t$  distribution  
<sup>146</sup> with 8 degrees of freedom, repeated 5000 times; the true skew is 0, and the true excess kurtosis is 1.5.  
<sup>147</sup> The distance between the largest and smallest estimates (indicated by the dotted red vertical lines), relative  
<sup>148</sup> to the distance between the 5th and 95th percentiles, shows the broad extent of extreme values that can  
<sup>149</sup> occur even with a large sample, especially for kurtosis.

<sup>150</sup> We therefore recommend nonparametric (NP) measures of skewness and kurtosis that are based on  
<sup>151</sup> quantiles and thus are less sensitive to a few extreme values. Let  $q_\alpha$  denote the  $\alpha$  quantile of a distribution  
<sup>152</sup> or sample (e.g.,  $q_{0.05}$  is the 5th percentile). For any  $0 < \alpha < 0.5$ , a quantile-based measure of skewness  
<sup>153</sup> is given by (McGillivray, 1986)

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

<sup>155</sup> NP Skewness measures the asymmetry between the tails of the distribution above and below the median.  
<sup>156</sup> The size of the upper tail can be measured (for any  $0 < \alpha < 0.5$ ) by  $\tau_U = q_{1-\alpha} - q_{0.5}$ ; for  $\alpha = 0.05$  this is the  
<sup>157</sup> difference between the 95th percentile and the median. The lower tail size is  $\tau_L = q_{0.5} - q_\alpha$ . The definition



**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`

158 above is equivalent to

$$159 \quad \text{NP Skewness} = \frac{\tau_U - \tau_L}{(\tau_U + \tau_L)}. \quad (2)$$

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

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

$$164 \quad \text{NP Kurtosis} = \frac{q_{1-\alpha} - q_\alpha}{q_{0.75} - q_{0.25}}. \quad (3)$$

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

170 Figure 2C,D illustrate how, applied to the same simulated samples, the nonparametric measures  
 171 produce a smaller fraction of highly inaccurate estimates caused by a few extreme values. Also note that,  
 172 in contrast to the moment-based measures, numerically small values of the nonparametric measures (e.g., 0.1  
 173 or 0.2) should not be disregarded, because both measures are scaled so that a value of 1 indicates extremely  
 174 large departures from a Gaussian distribution.

175 Using quantile-based measures carries the added value that quantile regression can be used to estimate  
 176 how they vary with initial size or expected future size. In the examples below, we use the **qgam** package  
 177 (Fasiolo *et al.*, 2020) to fit spline quantile regression models, which accommodate nonlinear size-dependence  
 178 in skewness and kurtosis. One risk of spline regression is that fitted quantiles may be excessively “wiggly”  
 179 without constraints on their complexity; with realistic amounts of data, we can hope to estimate broad trends  
 180 in distribution shape, but not fine-scale variation. In the examples below, we limit complexity by fitting  
 181 splines with  $k=4$  basis functions unless otherwise noted. Parametric quantile regression is also an option.

182 For consistency we also use quantile-based measures of mean and standard deviation when comparing  
 183 real and simulated data, and use quantile regression to visualize their trends. Specifically, following Wan  
 184 *et al.* (2014),

$$185 \text{NP Mean} = \frac{q_{0.25} + q_{0.5} + q_{0.75}}{3}, \quad \text{NP SD} = \frac{q_{0.75} - q_{0.25}}{1.35}. \quad (4)$$

186 **4 Case study: lichen, *Vulpicida pinastri***

187 We begin with a simple example where current size is the only predictor of future size. Growth data for  
188 the epiphytic lichen *Vulpicida pinastri* were analyzed first by Shriver et al. 2012 and again by Peterson  
189 et al. 2019 in their study of skewed growth distributions. We therefore had an *a priori* expectation of  
190 deviation from normality. The data set includes 1,542 inter-annual transitions in thallus area ( $cm^2$ ) observed  
191 from 2004 to 2009 in Kennicott Valley, AK. Shriver et al. 2012 used a mixture distribution that separated  
192 “normal growth or shrinkage” from “extreme shrinkage”. We aimed to fit a single growth model that could  
193 realistically accommodate both types of size transition without requiring *ad hoc* decisions about which  
194 observations of shrinkage were “extreme” or not.

195 With initial size as the only predictor, a convenient way to fit a Gaussian model with non-constant  
196 variance is the `gam` function in **mgcv** library (Wood, 2017) using the `gaulss` family. Following a bit  
197 of model selection, we fit the mean and standard deviation of future size as second-order polynomials  
198 of current size<sup>2</sup>, then calculated the scaled residuals from the fitted mean and standard deviation. Here,  
199 the first argument to `gam()` is a two-element list that defines the linear predictors for mean and sd:

```
200 # d is the data frame; t0,t1 are initial & final thallus area, respectively  
201 fitGAU <- gam(list(t1~t0 + I(t0^2), ~t0 + I(t0^2)), data=d, family=gaulss())  
202 d$fitted_mean = predict(fitGAU, type="response")[,1]  
203 d$fitted_sd <- 1/predict(fitGAU, type="response")[,2]  
204 d$scaledResids=residuals(fitGAU, type="response")/d$fitted_sd
```

205 The data and fitted mean and standard deviation are shown in Fig. 3A, and the corresponding diagnostic  
206 plots are in Fig. 4A,B. Our diagnostic plots are similar to plots made by R’s `plot.lm` function, except  
207 that we use spline regression to allow data-driven choice of curve smoothness, and use absolute residuals

---

<sup>2</sup>`gam()` is most commonly used to fit smooth splines (`s()`) for predictor variables, but it can also fit parametric regressions.

208 (rather than their square roots) so that the standard deviation of the regression curve is on the same scale  
209 as the residuals. The spline curves are not exactly flat – their standard deviations, given above each panel,  
210 are positive – but the trends are much too small to be worth fixing.

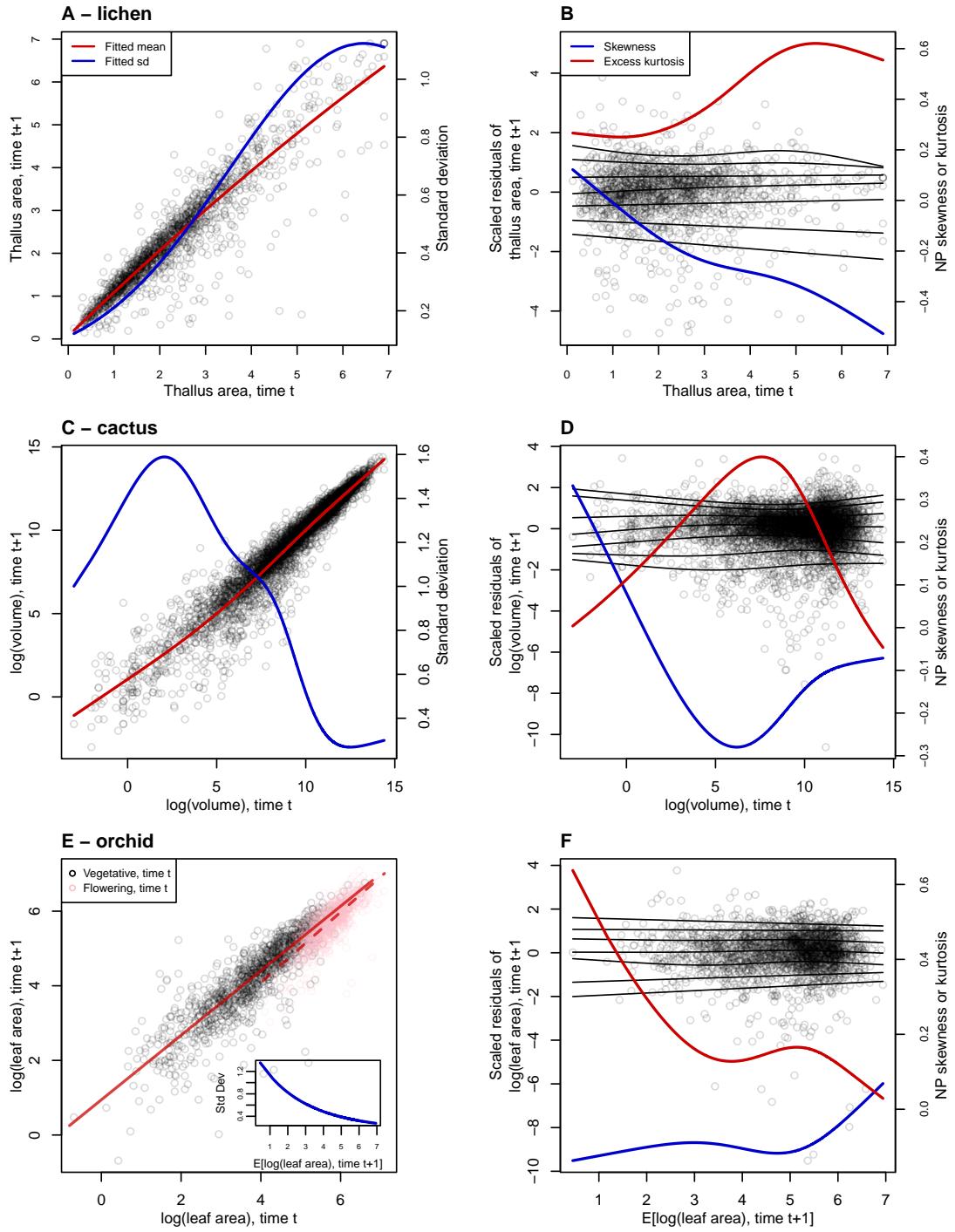
211 Quantile regression on the scaled residuals generates the skewness and kurtosis diagnostics shown  
212 in Fig. 3B. As expected based on previous analyses, the graphical analysis of the standardized residuals  
213 indicates negative skew, especially at larger sizes (Fig. 3B). We also find positive excess kurtosis for all sizes.

214 We turned to the Johnson's *S-U* (JSU) distribution for improvement. The JSU is a four-parameter  
215 leptokurtic distribution allowing positive or negative skew, with the convenient property that its location  
216 and scale parameters `mu` and `sigma` are the mean and standard deviation, respectively, which greatly  
217 facilitates the transition from a pilot Gaussian model. JSU is not available in any standard linear or additive  
218 modeling packages, to our knowledge. But that is not a barrier because we can write a likelihood function  
219 using the `dJSU()` function in the **gamlss.dist** package. Following the best-fit Gaussian model, we defined  
220 `mu` and `sigma` of the JSU as quadratic polynomials of initial size and, based on Fig. 3B) we define  
221 the skewness parameter `nu` as a linear function of size and kurtosis parameter `tau` as a positive constant.  
222 The likelihood function therefore has nine parameters to estimate. We fit the model using the **maxLik**  
223 package<sup>3</sup> with starting coefficient values for `mu` and `sigma` based on the pilot Gaussian model:

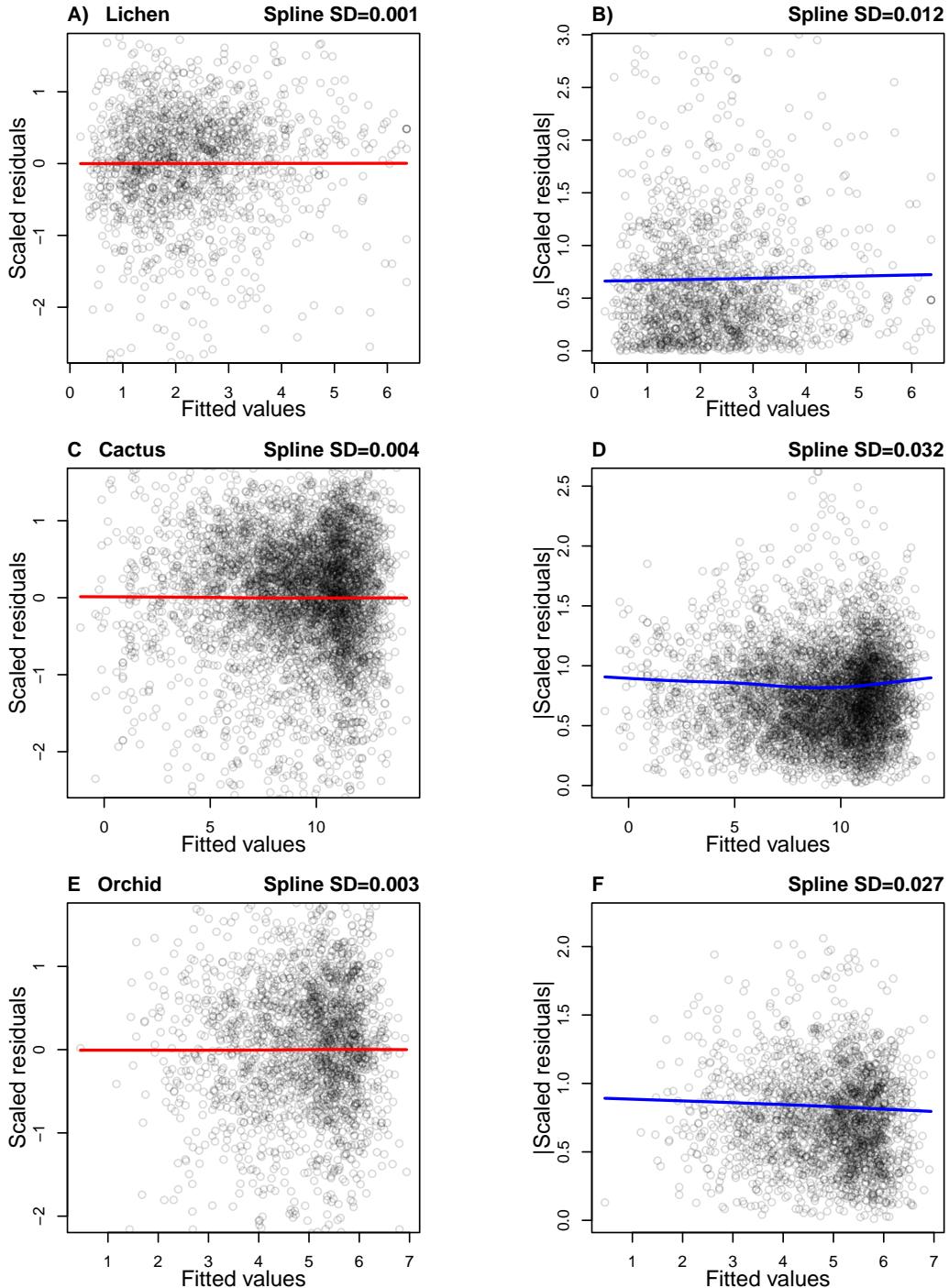
```
224 ## define function that returns the JSU negative log-likelihood  
225 LogLikJSU=function(pars){  
226   dJSU(t1, mu=pars[1]+pars[2]*t0+pars[3]*t0^2,  
227   sigma=exp(pars[4]+pars[5]*t0+pars[6]*t0^2),  
228   nu = pars[7]+pars[8]*t0, tau = exp(pars[9]), log=TRUE)  
229 }
```

---

<sup>3</sup>We chose **maxLik** because it offers the BHHH optimization method, which works well for non-Gaussian likelihoods in our experience.



**Figure 3:** Best Gaussian models and diagnostics of standardized residuals for lichen (*Vulpicida pinastri*) **A,B**, cactus (*Cylindriopuntia imbricata*) **C,D**, and orchid (*Orchis purpurea*) **E,F** case studies. **A,C**, fitted mean (red) and standard deviation (blue) of size at time  $t+1$  conditional on initial size at time  $t$ . **E**, fitted means for plants that were vegetative (solid line) or flowering (dashed line) at the start of the census interval and standard deviation as a function of the fitted mean (inset). **B,D,F** Quantile regressions of scaled residuals (lines show 5%, 10%, 25%, 50%, 75%, 90%, and 95% quantiles) and non-parametric measures of skewness (blue) and excess kurtosis (red) derived from them. In **B,D** scaled residuals are shown with respect to initial size and in **F** they are shown with respect to fitted values. Figure made by script `crosspp_growth.R`.



**Figure 4:** Diagnostic plot for trends in the mean (left column) or variance (right column) of scaled residuals from a pilot Gaussian model, for the lichen (*Vulpicida pinastri*) **A,B**, cactus *Cylindropuntia imbricata* **C,D**, and orchid *Orchis purpurea* **E,F** case studies. In **A,C,E** the standardized residuals are plotted, and in **B,D,F** the absolute values of standardized residuals, as functions of fitted mean subsequent size values. The solid curves are cubic splines (R function `smooth.spline`) fitted by generalized cross-validation with a modest over-penalization of model degrees of freedom to prevent overfitting (`penalty=1.4` as recommended by Gu (2013)). The numbers appearing above each panel are the standard deviation of the values on the spline regression curve, evaluated at all of the fitted values. Figure made by script `crossssp_diagnose_pilot.R`.

```

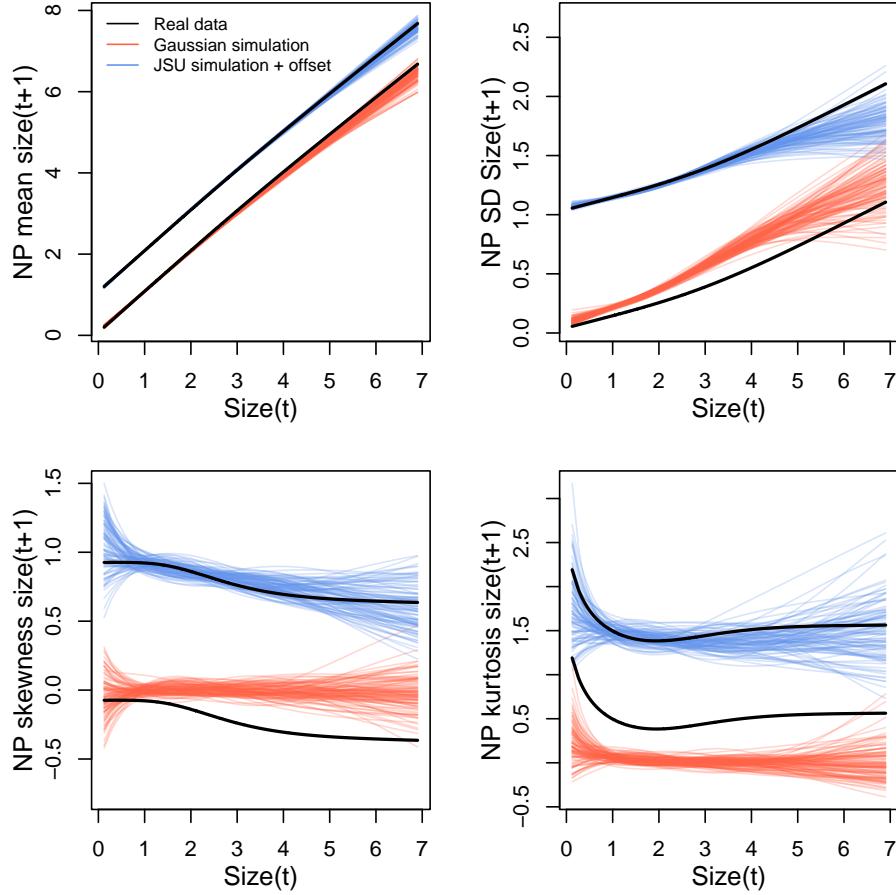
230 ## starting parameter values
231 p0<-c(coef(fitGAU)[1:6],0,0,0)
232 ## fit with maxlik, adding some noise to starting values
233 outJSU=maxLik(logLik=LogLikJSU,start=p0*exp(0.2*rnorm(length(p0))),
234 method="BHHH",control=list(iterlim=5000,printLevel=2),finalHessian=FALSE);

```

235 Simulating data from the fitted JSU model indicates a compelling improvement over the best Gaussian  
236 model, not only in skewness and kurtosis (Fig. 5C-D) but also the nonparametric standard deviation (5B).  
237 Note, in these data simulation figures Gaussian and non-Gaussian data are offset by an arbitrary amount  
238 to more easily visualize their correspondence to the real data (black lines in Fig. 5).

239 To understand the practical consequences of improved growth modeling, we assembled the remainder  
240 of the lichen IPM following Shriver et al. 2012. The asymptotic population growth rate  $\lambda$  based on Gaussian  
241 growth differs from the JSU growth model by about 1% annual population growth (Table 1), in line with  
242 results of Peterson et al. 2019. However, even this modest difference can lead to biased estimates of extinction  
243 risk from the Gaussian model, particularly over longer time horizons (Fig. 6). We also explored differences in  
244 other life history metrics (Table 1) using functions from Hernández *et al.* (2024). For example, the JSU growth  
245 model predicts values for mean lifespan, mean lifetime reproductive success, and generation time that are 15–  
246 25% lower than the Gaussian growth model. In this case study, properly modeling non-normal size transitions  
247 – which was easy to do with a few extra lines of code – can influence ecological inferences, at least based on  
248 point estimates. However, Table 1 also provides bias-corrected, bootstrapped confidence intervals (Diciccio &  
249 Efron, 1996), and these are heavily overlapping between the Gaussian and JSU models for all life history traits,  
250 suggesting that effects of “improved” growth modeling are small relative to our uncertainty in model parameters.

251 One could argue that this example was a convenient “straw man” to disqualify Gaussian growth,  
252 because it was recognized by the original and subsequent analysts that size transitions are strongly skewed



**Figure 5:** 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. Colored lines show 100 simulated data sets from the fitted Gaussian (red) or JSU (blue) growth models. Thick black line shows the real data. Gaussian and JSU data are offset by one unit and the real data line is duplicated with a one-unit offset for ease of visualization. Figure made by script `Vuplicida_IPMs.R`.

253 (Peterson *et al.*, 2019; Shriver *et al.*, 2012). In all remaining case studies, including those in the Supporting

254 Information, we re-examine growth data that were modeled as Gaussian in the original published analysis.

## 255 5 Case study: tree cholla cactus, *Cylindriopuntia imbricata*

256 The next case study, focused on the tree cholla cactus *Cylindriopuntia imbricata* at the Sevilleta Long-Term

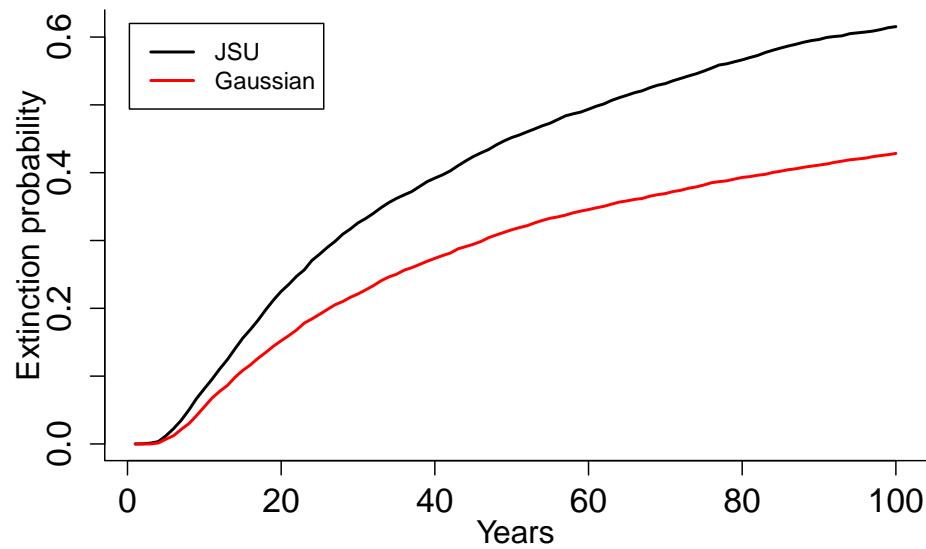
257 Ecological Research site in central New Mexico, adds a new feature to the simple size-dependent regressions

258 in the previous study: random effects associated with temporal (year) and spatial (plot) environmental

259 heterogeneity. This long-term study was initiated in 2004 and different subsets of the data have been analyzed

**Table 1:** Life history attributes derived from IPM kernels that included Gaussian or “improved” growth sub-models for six case studies. The improved distributions were JSU (lichen, creosote), SHASH (cactus, pike, coral), and skewed  $t$  (orchid). Pike, creosote and coral case studies are presented in the Supporting Information. The original coral case study assumed an open population with constant recruitment from a large source region, so some life history attributes cannot be computed from the published model. Values in parenthesis are 95% bootstrap confidence intervals, specifically the bias-corrected (BC) bootstrap confidence intervals Diciccio & Efron (1996). Table can be reproduced from scripts `crosspp_growth.R`, `Vulpicida_boot.R`, `Akumal_corals_boot.R`.

Species	Growth model	$\lambda$	Lifespan	Lifetime reproductive output	Age at reproduction	Generation time
Lichen ( <i>V. pinastri</i> )	Gaussian	1.01 (0.99, 1.04)	6.4 (3.6, 11.1)	1.4 (0.5, 3.1)	6.5 (5.7, 7.3)	40.8 (30.5, 57.4)
	Improved	1.00 (0.98, 1.03)	5.4 (3.1, 9.7)	1 (0.4, 2.4)	6.4 (5.4, 7.3)	36.6 (27.5, 48.6)
Cactus ( <i>C. imbricata</i> )	Gaussian	0.994 (0.99, 0.996)	6.11 (3.66, 8.63)	21.8 (8.27, 49.4)	17.6 (1.75, 22.7)	189 (131, 266)
	Improved	0.993 (0.991, 0.998)	5.38 (3.34, 16.3)	13.4 (5.72, 251)	20.3 (1.21, 22.2)	179 (133, 298)
Orchid ( <i>O. purpurea</i> )	Gaussian	1.09 (1.08, 1.1)	1.08 (1.06, 1.11)	20.0 (12.6, 31.0)	5.07 (4.78, 5.31)	104 (73.1, 150)
	Improved	1.09 (1.08, 1.1)	1.08 (1.06, 1.1)	19.3 (12.0, 29.9)	5.03 (4.75, 5.3)	100.7 (71.0, 145.0)
Pike ( <i>E. Lucius</i> )	Gaussian	1.62 (1.35, 1.89)	1.2 (1.09, 1.35)	5.75 (2.9, 9.7)	1.09 (1.03, 1.18)	4.96 (4.26, 5.84)
	Improved	1.62 (1.35, 1.88)	1.2 (1.09, 1.35)	5.76 (2.91, 9.73)	1.09 (1.03, 1.18)	4.94 (4.30, 5.84)
Creosote ( <i>L. tridentata</i> )	Gaussian	1.033 (1.029, 1.04)	$4.52 \times 10^6$ ( $2.14 \times 10^5$ , $1.82 \times 10^8$ )	$3.19 \times 10^5$ (1.27 $\times 10^4$ , $1.24 \times 10^7$ )	32.7 (29.2, 36.0)	$5.27 \times 10^6$ (2.50 $\times 10^5$ , $1.95 \times 10^8$ )
	Improved	1.034 (1.03, 1.04)	$3.26 \times 10^5$ ( $1.98 \times 10^3$ , $1.66 \times 10^7$ )	$2.31 \times 10^4$ (5.83 $\times 10^2$ , $1.27 \times 10^6$ )	32.8 (29.3, 36.0)	$3.7 \times 10^5$ (2.63 $\times 10^3$ , $1.93 \times 10^7$ )
Coral ( <i>G. ventalina</i> )	Gaussian	—	17.3 (11.9, 24.3)	—	10.5 (9.3, 11.8)	31.6 (28.3, 36.7)
	Improved	—	17.5 (12.1, 24.3)	—	10.7 (9.4, 12.2)	30.9 (27.4, 35.3)



**Figure 6:** Extinction risk estimated from individual-based simulation of IPMs based on Gaussian and Johnson’s S-U (JSU) growth distributions. Figure made by script `Vulpicida_IPMs.R`.

260 in various IPM studies, all using Gaussian growth kernels (Compagnoni *et al.*, 2016; Czachura & Miller, 2020;  
261 Elderd & Miller, 2016; Miller *et al.*, 2009; Ohm & Miller, 2014). In fact, Elderd and Miller 2016 presented  
262 a Gaussian growth model as an example of a well fit growth function, based on an overall distribution of  
263 residuals that appeared Gaussian and posterior predictive checks (PPCs) of a Bayesian model that suggested  
264 consistency between the real data and data simulated from the fitted model (Fig. 4 in (Elderd & Miller,  
265 2016)). While PPCs and the associated “Bayesian P-value” are popular diagnostic tools, they are often too  
266 conservative (Conn *et al.*, 2018; Zhang, 2014), failing to reject marginally bad models even though they are  
267 very effective in rejecting terrible models. The choice of discrepancy function (the statistic used to compare  
268 real and simulated data) can also be limiting: in our previous work, we used a discrepancy function focused  
269 on variance (the sum of squared residuals), creating a blind spot for poor modeling of higher moments.

270 The data includes 4844 size transition observations from 929 individuals spanning 13 transition years  
271 (2004–2018) and 11 spatial replicates (three spatial blocks in years 2004–2008 and eight 30m-by-30m plots  
272 in years 2009–2018). The data are provided in Miller (2020). Following previous studies, we quantified  
273 size as the natural logarithm of plant volume ( $cm^3$ ), derived from height and width measurements.

274 We begin growth modeling, as above, with a generalized additive model with the mean and standard  
275 deviation of size in year  $t+1$  modeled as smooth function of size in year  $t$ , with random intercepts for  
276 year and plot and assuming normally-distributed residuals:

```
277 # t0 and t1 are initial and final log(volume), respectively  
278 fitGAU <- gam(list(t1 ~ s(t0,k=4) + s(plot,bs="re") + s(year,bs="re"),  
279 ~ s(t0,k=6)), data=caactus, family=gaulss())
```

280 Note that here we fitted the standard deviation function with  $k=6$  basis functions rather than our default  
281 of  $k=4$  because, in a preliminary analysis, we found a moderate variance trend in the standardized residuals

282 using  $k = 4$ , suggesting a need for greater flexibility. With  $k = 6$ , spline regression detected essentially  
283 no trend in the mean of the resulting standardized residuals (Fig. 4C,D).

284 The growth variance is estimated to peak at small to medium sizes (Fig. 3C). The standardized  
285 residuals show clear signals of negative skew and positive excess kurtosis across most of the size distribution,  
286 but strongest in the middle (Fig. 3D). We therefore need a distribution family allowing negative skew  
287 and positive excess kurtosis, both of which may be negligible at some sizes. We first tried Johnson's  $S_U$  and  
288 then the skewed  $t$  distributions, which provided some improvements but there were still visible discrepancies  
289 between simulated and real data. We next turned to the SHASH distribution, which allows a greater range  
290 of kurtosis for a given amount of skew, and vice versa (Jones & Pewsey (2009); Supporting Information  
291 S.1). This flexibility proved necessary to generate simulated data that compared favorably to the real data,  
292 so we proceeded with the SHASH. Conveniently, SHASH is available as an **mgcv** family, allowing for  
293 flexible size-dependence in skewness and kurtosis without having to select specific size-dependent functions.

294 Here, the first argument to `gam()` is now a four-element list specifying the linear predictors for  
295 the four parameters of the SHASH distribution.

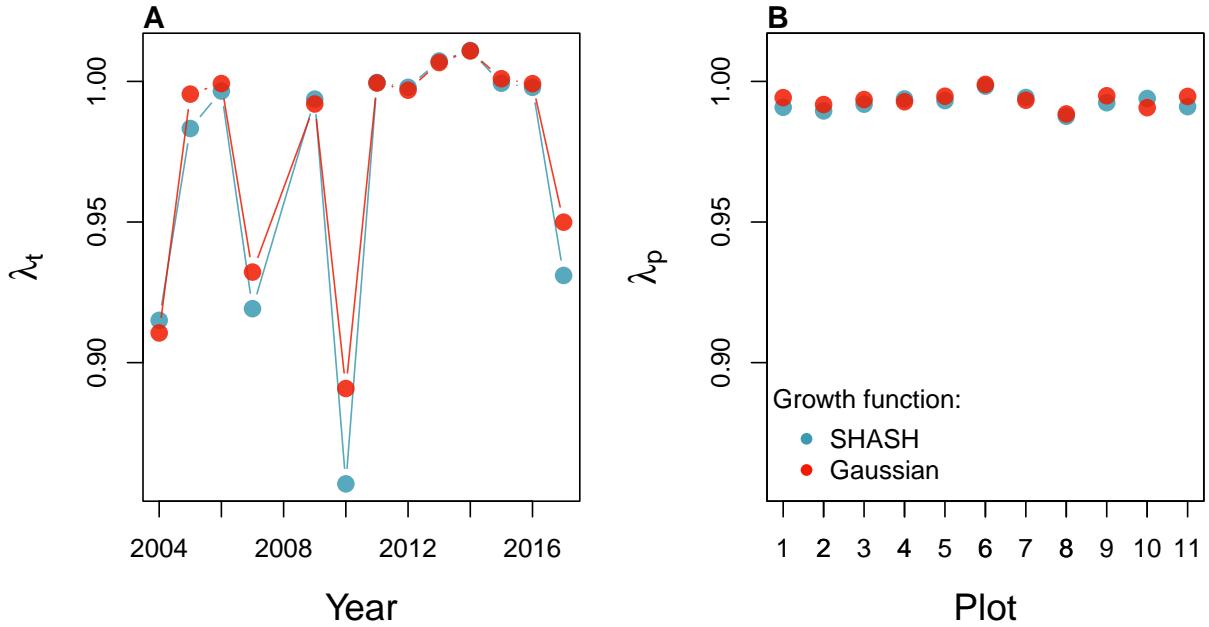
```
296 fit_shash <- gam(list(t1 ~ s(t0,k=4) +  
297   s(plot,bs="re") + s(year_t,bs="re"), # location  
298   ~ s(t0,k=4), # log-scale  
299   ~ s(t0,k=4), # skewness  
300   ~ s(t0,k=4)), # log-kurtosis  
301   data = cactus, family = shash,optimizer = "efs")
```

302 Data simulated from the SHASH model compared favorably to the real data (Fig. S-11). Similar to  
303 the lichen case study, we see that correctly modeling skewness and kurtosis improved estimation of the  
304 nonparametric mean and standard deviation (Fig. S-11A,B), yielding a growth model that is truer to the data.

We next explored how improved growth modeling influenced IPM results. The  $\lambda$  values predicted by Gaussian and SHASH growth functions, corresponding to the average plot and year, were nearly identical (Table 1) but we could also leverage structure of the study design to quantify demographic variance associated with temporal and spatial heterogeneity. We used the fitted random effects from the vital rate models to estimate the asymptotic growth rate for each year ( $\lambda_t$ ), centered on the average plot, and for each plot ( $\lambda_p$ ), centered on the average year. Estimates of  $\lambda_t$  from the Gaussian growth model were often greater than estimates from the SHASH growth model, particularly in some of the harshest years (Fig. 7A), and therefore the Gaussian model predicted lower temporal variance in fitness ( $SD(\lambda_{t(Gaussian)}) = 0.04$ ,  $SD(\lambda_{t(SHASH)}) = 0.048$ ). Plot-to-plot variation was more similar between the two models ( $SD(\lambda_{p(Gaussian)}) = 0.0026$ ,  $SD(\lambda_{p(SHASH)}) = 0.0028$ ), although spatial variation in fitness was much lower than temporal variation (Fig. 7B). The difference in temporal variance would suggest that Gaussian growth modeling would predict a higher stochastic growth rate  $\lambda_S$ , because temporal variance has a negative effect on  $\lambda_S$ . However, the stochastic growth rate from the Gaussian growth model ( $\lambda_S = 0.992$ ) was nearly identical to that of the SHASH growth model ( $\lambda_S = 0.991$ ). This is likely because temporal fluctuations in vital rates, which is where the SHASH growth model would make a difference, have a weaker influence on  $\lambda_S$  than the temporal fluctuations in size structure that they generate (Compagnoni *et al.*, 2016; Ellis & Crone, 2013). The SHASH and Gaussian growth models predicted small differences in other life history traits but, as in the lichen case study, these differences were small relative to the uncertainty captured by bootstrapped confidence intervals (Table 1). Thus, in this case study, modeling non-Gaussian size transitions with a Gaussian growth model may or may not influence IPM results depending on the target of the analysis.

## 6 Case study: lady orchid, *Orchis purpurea*

Our final case study examines selection on life history strategies in the lady orchid *Orchis purpurea*. In a prior study, Miller *et al.* 2012 analyzed how costs of reproduction (flowering or not in year  $t$ ) affected



**Figure 7:** Temporal (A) and spatial (B) heterogeneity in fitness for the tree cholla cactus (*Cylindropuntia imbricata*) predicted by IPMs using Gaussian or SHASH growth models. Figure made by script `cactus_growth_modeling_qgam.R`.

328 growth from year  $t$  to  $t+1$ . The two growth kernels for flowering and non-flowering were then used in  
 329 an IPM to quantify the optimal flowering size that balances the benefits of waiting to flower at larger sizes  
 330 against the greater risk of death before flowering. The original study assumed Gaussian size transitions  
 331 with non-constant variance depending on initial size. Here we re-visit that analysis to derive improved  
 332 growth kernels. We use this case study to illustrate several new elements and challenges, including modeling  
 333 skewness and kurtosis as functions of expected future size.

334 The data, originated by Dr. Hans Jacquemyn, come from 368 plants in a Belgian population censused  
 335 annually from 2003 through 2011. Here we use data only from the “light” habitat in the original study.  
 336 We used the natural logarithm of total leaf area as the size variable in the IPM.

337 As a variation on software, we fitted the pilot Gaussian model using the `lmer` function in the **lme4**  
 338 package, as in the original study. We fit three candidate linear models that included fixed effects of size  
 339 in year  $t$  (model 1), additive effects of size and flowering status in year  $t$  (model 2), or an interaction between  
 340 size and flowering (model 3), all including random intercepts for year. The interaction model was strongly

341 favored ( $\Delta AIC = 10.5$ ). Unlike our previous case studies, here we have multiple fixed effects (initial size  
342 and flowering status) that may influence the variance of future size. In cases such as this it is convenient  
343 to model variance as a function of expected future size, rather than initial size as we did with the lichens  
344 and cacti. The expected (or “fitted”) values reflect the combined influence of all fixed and random effects,  
345 and therefore implicitly account for multiple sources of variation in the variance.

346 Models where error variance is a function of fitted values cannot be fitted directly with `lme4` (nor in  
347 the `mgcv` functions for generalized additive models). But it can still be done with `lmer` through an iterative  
348 re-weighting approach, as follows. In `lmer`, weights  $w_i$  can be used to indicate that the observations  $y_i$   
349 have error variance proportional to  $1/w_i^2$ . The iterative steps are as follows, and code that executes these  
350 steps is in `orchid_growth_modeling.R`.

- 351 1. Fit the expected value assuming Gaussian-distributed residuals with constant variance.  
352 2. Fit the standard deviation of the residuals as a function of the corresponding fitted value.  
353 3. Re-fit the model, with weights equal to the inverse of the standard deviation estimated in step 2.

354 We iterated steps 2 and 3 until the root mean square change in weights was below  $10^{-6}$ . This is not elegant,  
355 but it works and converges quickly. In step 2, we modeled the log of the standard deviation (because standard  
356 deviations cannot be negative) as a quadratic polynomial in the fitted mean. In exploratory analyses we found  
357 that the quadratic term was necessary to fit the standard deviation. We did this for all candidate models and,  
358 for a fair AIC comparison, we then re-fit all candidate models with the weights estimated from the top model.

359 The updated model selection continued to favor the size  $\times$  flowering interaction model (3), but now  
360 with a weaker improvement over the next-best model ( $\Delta AIC = 6.7$ ). The fitted mean (a function of initial  
361 size and flowering status) and fitted standard deviation (a function of the fitted mean) are shown in Fig.  
362 3E. Spline regression found no trend in the mean of the resulting standardized residuals, and only small  
363 variation in the variance (Fig. 4E,F).

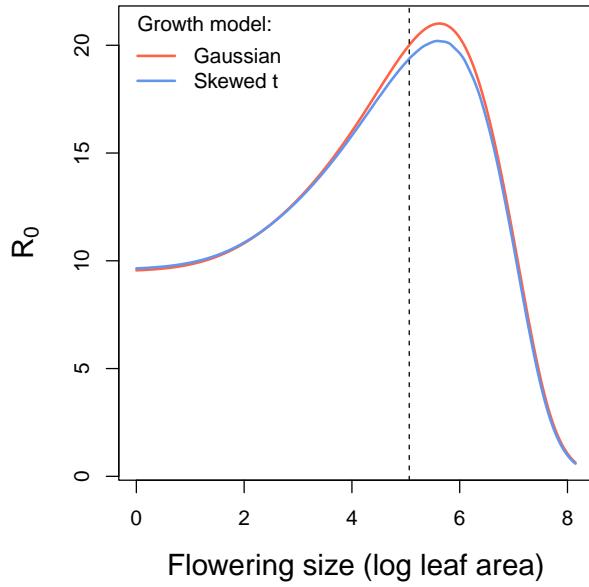
364       The best Gaussian model indicated a growth cost associated with flowering at the start of the census  
365      interval and a decline in growth variance with increasing expected values (Fig. 3E). The standardized  
366      residuals indicated negative skewness (10–20% difference in tail weight) and excess kurtosis (10–40% fatter  
367      than Gaussian) across much of the size distribution but both negligible at large expected sizes (Fig. 3F).

368       As possible improvements, we explored the skewed *t* and JSU distributions, both leptokurtic  
369      distributions with flexible skewness. Based on comparisons between real and simulated data we were  
370      happier with the skewed *t*, which we fit with a custom likelihood function similar to the JSU growth model  
371      for the lichen data. However, rather than re-fitting all parameters of the skewed *t* model, as we did with  
372      the lichen JSU, we built a “hybrid” likelihood function that uses the fitted mean and standard deviation from  
373      the best Gaussian model, and estimates parameters that control skewness and kurtosis as linear functions  
374      of expected future size. This is easy because the **gamlss.dist** package provides a parameterization of the  
375      skewed *t* in which the location parameter  $\mu$  is the mean and scale parameter  $\sigma$  is the standard deviation  
376      (Rigby *et al.*, 2019). The hybrid likelihood looks like this:

```
377 ## GAU_fitted and GAU_sd are mean & standard deviation from the best Gaussian.  
378 SSTLogLik=function(pars){  
379   dSST(log_area_t1,  
380   mu=GAU_fitted, sigma=GAU_sd,  
381   nu = exp(pars[1] + pars[2]*GAU_fitted),  
382   tau = exp(pars[3] + pars[4]*GAU_fitted)+2, log=TRUE)  
383 }  
384 p0<-c(0,0,0,0) ## default starting parameters  
385 SSTout=maxLik(logLik=SSTLogLik,start=p0) ## fit with maxLik
```

386 Based on diagnostics of the standardized residuals, parameters that control skewness and kurtosis are defined as  
387 linear functions of the mean (note that the `tau` parameter uses a  $\log(x-2)$  link function). This approach relies  
388 on the robustness of fitted Gaussian models to deviations from normality, which implies that the fitted mean  
389 and variance from a Gaussian model are good approximations for the mean and variance of the corresponding  
390 non-Gaussian model. If one is skeptical of this approach, it is possible to simultaneously re-fit all parameters of  
391 the skewed  $t$ . However, recall that unlike the lichen case study, the pilot Gaussian model here includes random  
392 year effects, and the expected values getting passed into dSST account for this source of variation. Estimating  
393 random effects “from scratch” with a custom likelihood model is possible (we provide guidance on doing this  
394 with a “shrinkage” approach, in Supporting Information S.2), but generally should not be necessary. Instead, a  
395 key advantage of the hybrid approach is retention of the fitted random effects and associated variance compo-  
396 nents, which get shuttled from the Gaussian model into the non-Gaussian model without any fuss (though it was  
397 critical to use a parameterization of the skewed  $t$  for which `mu` is the mean and `sigma` is the standard deviation).  
398 And, if this approach does not “work” (i.e., deviations from normality biased the fitted values of the Gaussian  
399 model) one would quickly find out when comparing simulated with real data. In this case, size transition  
400 data simulated from this model corresponded favorably to the real data, much better than the pilot Gaussian  
401 model, including improvements in the standard deviation, skewness, and kurtosis of future size (Fig. S-12).

402 Finally, we used the improved growth model to revisit key results of the original study. Miller et  
403 al. (2012) used the orchid IPM to estimate the evolutionarily stable strategy (ESS) as the mean size at  
404 flowering that maximizes lifetime reproductive success ( $R_0$ ), given the constraint that flowering when  
405 small reduces growth and thus elevates mortality risk. Repeating that analysis here, we found that improved  
406 growth modeling has virtually no influence on predictions for optimal life history strategies (Fig. 8). ESS  
407 flowering sizes were nearly identical between IPMs with Gaussian vs skewed  $t$  growth models, and both  
408 aligned well with the observed mean flowering size (dashed vertical line in Fig. 8). **Similarly, there were**



**Figure 8:** Orchid life history results from IPMs using Gaussian or skewed  $t$  growth models. Lifetime reproductive success ( $R_0$ ) is shown as a function of mean size of flowering. Dashed vertical line shows the observed mean flowering size.

409 very small differences between growth functions in other metrics of orchid life history and, again, these  
 410 differences were overwhelmed by uncertainty associated with parameter estimation (Table 1).

## 411 7 Discussion

412 Much of the appeal of IPMs has stemmed from their embrace of continuous size structure through regression-  
 413 based approaches, and the potentially complex fixed- and random-effect structures that those approaches allow.  
 414 Using familiar statistical tools and with relatively few parameters to estimate, IPM users can incorporate  
 415 important sources of variation in demography and interrogate their influence on ecological and evolutionary  
 416 dynamics. With this opportunity comes the burden of getting it right: an IPM is only as good as the statistical  
 417 sub-models for the underlying data. The growth sub-model is the trickiest part because it defines a distribution  
 418 of future size conditional on current size. Distributions have many properties – “moments” – and a good  
 419 growth model should recapitulate the properties of real size transitions. The default assumption of Normally  
 420 distributed size transitions, employed overwhelmingly across 20+ years of IPM studies, is an arbitrary historical  
 421 precedent. In our case studies and, we suspect, more broadly, skewness and excess kurtosis were common

422 features of size transitions. Our most important message is that the assumption of normally-distributed size  
423 transitions can easily be abandoned, and a more inquisitive process of growth modeling should take its place.

424 We have attempted to lay out what that process should look like, emphasizing visual diagnostics  
425 to characterize how data deviate from Gaussian. One implication of relying on visual diagnostics is that  
426 goodness of fit is in the eye of the beholder. This empowers IPM users to make informed choices, but it is not  
427 very prescriptive; we have not suggested any hard rules for choosing among distributions, only that a good  
428 growth model should generate data that look like the real thing. Alternatively, model selection could be used  
429 to identify best-fitting growth distributions and best-fitting functions for higher moments. However, model  
430 selection among growth distributions with 3-5 parameters, each of which may be functions of multiple state  
431 variables or fitted values, can quickly explode in complexity, and we are not convinced it is worth the trouble.

432 Our work follows the important contribution of Peterson et al. 2019, who were similarly motivated  
433 by inadequacy of the Gaussian model but arrived at different recommendations. These authors developed  
434 a creative approach in which size data are transformed onto a [0,1] scale and size transitions on that scale  
435 are modeled using beta regression. The beta distribution can accommodate positive, negative, or zero skew,  
436 potentially varying with size, so the Peterson et al. approach is a flexible option for skewed growth data.  
437 However, beta regression also has some limitations: common beta regression packages do not fit random  
438 effects (e.g., **betareg** (Cribari-Neto & Zeileis, 2010)) or do not do so reliably (in our experience **gamlss**  
439 regressions with random effects are numerically unstable); and the two-parameter beta distribution does not  
440 allow skewness and kurtosis to be fitted independently. Additionally, the initial transformation onto [0,1] scale  
441 requires estimating extreme quantiles of the growth distribution (e.g., 0.01 and 0.99) as a function of initial size.  
442 In our experience those quantile estimates can be very sensitive to how size-dependence is modeled, and model  
443 selection is challenging for extreme quantiles where data are (by definition) very sparse. Rather than picking  
444 one distribution as a new default, users can leverage the vast arsenal of continuous probability distributions  
445 – all at one's fingertips with a few lines of code – so that the data and their particular deviations from

446 normality can guide the choice of a better distribution. It is also possible to use mixtures of multiple growth  
447 distributions, as done by Shriver et al. 2012 to model “normal” and “extreme” types of lichen shrinkage. In  
448 re-analyzing that data set, we found that a single, flexible distribution (Johnson’s *S-U*) could recapitulate the  
449 observed size transitions, but there may be other cases where mixture distributions are preferable or necessary.

450 In all of our case studies, non-Gaussian growth models always yielded more satisfying fits to size  
451 transition data than the Gaussian models published in those papers. However, to our relief, none of these  
452 re-analyses yielded a “gotcha” result that overturned results of the original study. In fact, in this small  
453 sampling of case studies, improved growth modeling had weak to modest effects on IPM results, similar in  
454 magnitude to the results of Peterson *et al.* (2019), and for most species and life history metrics these effects  
455 were overwhelmed by the uncertainty associated with parameter estimation (Table 1). For some case studies,  
456 one might argue that non-Gaussian modeling was not worth the trouble – only it was almost no trouble  
457 at all, and we could not have known whether or not a non-Gaussian model would have made a difference  
458 before fitting it. Even where Gaussian and improved growth models differed in IPM results, we do not know  
459 “true” values to compare them against, only that one describes the data better than another, and a closer  
460 match to the data does not necessarily translate to better predictive ability due to the bias-variance trade-off.

461 We caution against taking too much comfort in weak effects of “improved” growth modeling; in other  
462 scenarios the choice of growth distribution could be more consequential. We focused on life history metrics  
463 such as mean lifespan and mean lifetime reproductive success (Table 1). It is possible that higher moments  
464 of those traits (e.g., variance or skewness of lifespan and lifetime reproduction) are more sensitive to the  
465 tails of the growth distribution. It is also worth noting that most of our case studies focused on perennial  
466 life histories (perennial plants and lichens) characterized by relatively slow growth, heavy losses during  
467 recruitment, and high survival once established, and these species all had mean lifespans between one and six  
468 years and generation times on the order of decades. Life histories such as these may be relatively robust to  
469 subtle features of the growth kernel. In the Supporting Information we present three additional case studies

470 that broaden our life history coverage, including pike (*Esox lucius*), a fish with a generation time of four  
471 to five years and creosotebush (*Larrea tridentata*), a desert shrub that is virtually immortal once established.  
472 Life history metrics from these “fast” and “slow” populations were no more sensitive to improved growth  
473 modeling than those of the perennial plants and lichens (Table 1). More systematic comparative analyses  
474 may provide insight into which types of species and life histories are more likely to exhibit strong skewness  
475 and kurtosis, and which demographic quantities are more or less sensitive to these features of size transition.

476 Our case studies illustrate a diversity of software packages and computational approaches, to reflect the  
477 diversity of preferences and habits that the community of IPM analysts bring to their own problems. We like  
478 spline generalized additive models (gams) for their flexibility and for **mgcv**’s numerous options for distribution  
479 families and overall speed and reliability. However, there are some applications for which classical parametric  
480 regression would be preferable because the coefficients carry biological meaning. For example, regression  
481 coefficients may be targets of natural selection (Rees & Ellner, 2016) and may combine to influence traits  
482 of interest such as the expected size at flowering (e.g. in Fig. 8A), a function of the intercept and slope of the  
483 size-dependent flowering function (Metcalf *et al.*, 2003). Some potentially useful distributions are not available  
484 in linear modeling software packages, but that should not be a barrier to their use: as in several of our case  
485 studies, custom likelihood functions allow non-Gaussian models without sacrificing the complex, multi-level  
486 features that one might be accustomed to fitting in **lme4**, for example. Bayesian analysis may further broaden  
487 the options for non-Gaussian candidate distributions and may help estimate hard-to-fit parameters through the  
488 brute force of sampling algorithms. Bayesian analysis also provides a natural way to propagate uncertainty from  
489 vital rate sub-models to full model predictions (Elderd & Miller, 2016), as an alternative to our bootstrapping  
490 approach. However, as of this writing, most of the non-Gaussian distributions that we have discussed are  
491 not available in popular Bayesian software packages such as Stan or JAGS. While user-defined distributions  
492 can always be coded from scratch, this may be a significant technical barrier for many IPM analysts.

493 From the outset there have been concerns about “how well these methods [IPM growth kernels]  
494 can deal with different patterns of growth, stasis, and shrinkage” (Morris & Doak, 2002, p. 200), compared  
495 to “binning” methods that use observed transition frequencies between user-defined size classes as the  
496 transition probabilities in a (possibly large) matrix model (Doak *et al.*, 2021). The non-Gaussian models that  
497 we have considered here are not a panacea. For example, none of them allow bimodal growth, such as might  
498 occur if herbivore- or pathogen-attached individuals experience rapid tissue loss. When the shape of the  
499 growth distribution is nearly the same for all initial sizes, a nonparametric IPM growth kernel can be defined  
500 from a kernel density estimate for scaled residuals (Ellner *et al.*, 2016, p. 288). Outside that special situation,  
501 nonparametric approaches require choosing multiple smoothing parameters, which is very challenging.  
502 We are currently exploring whether “targeted learning” approaches developed for causal inference (van der  
503 Laan & Rose, 2011) can be used to circumvent smoothing parameter selection. Targeted learning starts  
504 with a pilot model and updates it iteratively to achieve unbiased estimates and valid confidence intervals  
505 for a particular “target” quantity, such as  $\lambda$  or mean lifespan. Preliminary results suggest that targeted  
506 learning with a deliberately under-smoothed pilot model works well for complex growth patterns (?). But  
507 nonparametric methods are data-hungry, so when departures from Gaussian are quantitative rather than  
508 qualitative, parametric modeling as developed here will make more efficient use of limited data.

## 509 Conclusion

510 Gaussian-distributed size transitions are probably the exception in nature, not the rule, yet two decades  
511 of IPM studies have relied overwhelmingly on Gaussian growth models. Using tools not available when  
512 IPMs were first developed, it should often be possible now to make major improvements over a Gaussian  
513 model, without worrying about finding the “best” alternative. By generating predicted size transitions  
514 that are truer to the data, IPM analysts can narrow the gap between model and nature.

515 **Acknowledgements:** This research was supported by US NSF grants DEB-1933497 to SPE and  
516 DEB-1754468, 2208857, and 2225027 to TEXM. The Sevilleta LTER (source of the cactus and creosote  
517 case studies) is supported by DEB-1655499 and DEB-1748133. Giles Hooker gave us the very good  
518 idea to use quantile regression instead of binning to estimate trends in skewness and kurtosis. Ali Campbell  
519 and Jacob Moutouama provided helpful discussion and comments on the manuscript.

520 **Authorship statement:** All authors discussed all aspects of the research and contributed to developing  
521 methods, analyzing data, and writing and revising the paper.

522 **Conflict of interest statement:** The authors have none to declare.

## 523 Literature Cited

524 Adler, P.B., Kleinhesselink, A., Hooker, G., Taylor, J.B., Teller, B. & Ellner, S.P. (2019) Data from: Weak  
525 interspecific interactions in a sagebrush steppe? Conflicting evidence from observations and experiments.  
526 Dryad Data set. <https://doi.org/10.5061/dryad.96dn293>.

527 Anscombe, F.J. & Glynn, W.J. (1983) Distribution of the kurtosis statistic  $b_2$  for normal samples. *Biometrika*  
528 **70**, 227–234.

529 Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using lme4.  
530 *Journal of Statistical Software* **67**, 1–48.

531 Compagnoni, A., Bibian, A.J., Ochocki, B.M., Rogers, H.S., Schultz, E.L., Sneck, M.E., Elderd, B.D.,  
532 Iler, A.M., Inouye, D.W., Jacquemyn, H. *et al.* (2016) The effect of demographic correlations on the  
533 stochastic population dynamics of perennial plants. *Ecological Monographs* **86**, 480–494.

534 Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R. & Hooten, M.B. (2018) A guide to Bayesian model  
535 checking for ecologists. *Ecological Monographs* **88**, 526–542.

- 536 Cooch, E.G. & White, G.C. (2020, accessed 5/17/2020) *Program MARK - a 'gentle introduction'*. phidot.org.
- 537 Coulson, T. (2012) Integral projections models, their construction and use in posing hypotheses in ecology.
- 538        *Oikos* **121**, 1337–1350.
- 539 Cribari-Neto, F. & Zeileis, A. (2010) Beta regression in R. *Journal of statistical software* **34**, 1–24.
- 540 Crone, E.E. (2016) Contrasting effects of spatial heterogeneity and environmental stochasticity on population
- 541        dynamics of a perennial wildflower. *Journal of Ecology* **104**, 281–291.
- 542 Czachura, K. & Miller, T.E. (2020) Demographic back-casting reveals that subtle dimensions of climate
- 543        change have strong effects on population viability. *Journal of Ecology* **108**, 2557–2570.
- 544 D'Agostino, R.B. (1970) Transformation to normality of the null distribution of  $g_1$ . *Biometrika* **57**, 679–681.
- 545 Davis, C. (2015) *sgt: Skewed Generalized T Distribution Tree*. R package version 2.0.
- 546 DiCiccio, T.J & Efron, B. (1996) Bootstrap Confidence Intervals. *Statistical Science* **11**, 189–228.
- 547 Doak, D.F., Waddle, E., Langendorf, R.E., Louthan, A.M., Isabelle Chardon, N., Dibner, R.R., Keinath,
- 548        D.A., Lombardi, E., Steenbock, C., Shriver, R.K., Linares, C., Begoña Garcia, M., Funk, W.C., Fitzpatrick,
- 549        S.W., Morris, W.F. & DeMarche, M.L. (2021) A critical comparison of integral projection and matrix
- 550        projection models for demographic analysis. *Ecological Monographs* **91**, e01447.
- 551 Easterling, M.R., Ellner, S.P. & Dixon, P.M. (2000) Size-specific sensitivity: applying a new structured
- 552        population model. *Ecology* **81**, 694–708.
- 553 Elderd, B.D. & Miller, T.E. (2016) Quantifying demographic uncertainty: Bayesian methods for integral
- 554        projection models. *Ecological Monographs* **86**, 125–144.

- 555 Ellis, M.M. & Crone, E.E. (2013) The role of transient dynamics in stochastic population growth for nine  
556 perennial plants. *Ecology* **94**, 1681–1686.
- 557 Ellner, S.P., Adler, P.B., Childs, D.Z., Hooker, G., Miller, T.E. & Rees, M. (2022) A critical comparison of  
558 integral projection and matrix projection models for demographic analysis: Comment. *Ecology* **103**, e3605.
- 559 Ellner, S.P., Childs, D.Z. & Rees, M. (2016) *Data-driven Modeling of Structured Populations: A Practical*  
560 *Guide to the Integral Projection Model*. Springer, New York.
- 561 Fasiolo, M., Wood, S.N., Zaffran, M., Nedellec, R. & Goude, Y. (2020) qgam: Bayesian non-parametric  
562 quantile regression modelling in R. *arXiv preprint arXiv:2007.03303*.
- 563 Gu, C. (2013) *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New York, 2 edn.
- 564 Hadfield, J.D. *et al.* (2010) MCMC methods for multi-response generalized linear mixed models: the  
565 mcmcglmm r package. *Journal of Statistical Software* **33**, 1–22.
- 566 Hérault, B., Bachelot, B., Poorter, L., Rossi, V., Bongers, F., Chave, J., Paine, C.T., Wagner, F. & Baraloto,  
567 C. (2011) Functional traits shape ontogenetic growth trajectories of rain forest tree species. *Journal*  
568 *of ecology* **99**, 1431–1440.
- 569 Hernández, C.M., Ellner, S.P., Snyder, R.E. & Hooker, G. (2024) The natural history of luck: A synthesis  
570 study of structured population models. *Ecology Letters* pp. 0 – 1.
- 571 Komsta, L. & Novomestky, F. (2015) moments: Moments, cumulants, skewness, kurtosis and related  
572 tests. *R package version 0.14.1*.
- 573 Louthan, A.M., Keighron, M., Kiekebusch, E., Cayton, H., Terando, A. & Morris, W.F. (2022) Climate  
574 change weakens the impact of disturbance interval on the growth rate of natural populations of venus  
575 flytrap. *Ecological Monographs* **92**, e1528.

- 576 McGillivray, H. (1986) Skewness and asymmetry: measures and orderings. *Annals of Statistics* **14**, 994–1011.
- 577 Metcalf, C.J.E., Ellner, S.P., Childs, D.Z., Salguero-Gómez, R., Merow, C., McMahon, S.M., Jongejans, E. & Rees, M. (2015) Statistical modelling of annual variation for inference on stochastic population dynamics using Integral Projection Models. *Methods in Ecology and Evolution* **6**, 1007–1017.
- 580 Metcalf, J.C., Rose, K.E. & Rees, M. (2003) Evolutionary demography of monocarpic perennials. *Trends in Ecology & Evolution* **18**, 471–480.
- 582 Miller, T.E. (2020) Long-term study of tree cholla demography in the los pinos mountains, sevilleta national wildlife refuge. <https://doi.org/10.6073/pasta/dd06df3f950afe4a4642306182237d13>.
- 584 Miller, T.E., Louda, S.M., Rose, K.A. & Eckberg, J.O. (2009) Impacts of insect herbivory on cactus population dynamics: experimental demography across an environmental gradient. *Ecological Monographs* **79**, 155–172.
- 587 Miller, T.E., Williams, J.L., Jongejans, E., Brys, R. & Jacquemyn, H. (2012) Evolutionary demography of iteroparous plants: incorporating non-lethal costs of reproduction into integral projection models. *Proceedings of the Royal Society B: Biological Sciences* **279**, 2831–2840.
- 590 Morris, W.F. & Doak, D.F. (2002) *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis*. Sinauer Associates, Sunderland, Mass.
- 592 Ohm, J.R. & Miller, T.E. (2014) Balancing anti-herbivore benefits and anti-pollinator costs of defensive mutualists. *Ecology* **95**, 2924–2935.
- 594 Ozgul, A., Childs, D.Z., Oli, M.K., Armitage, K.B., Blumstein, D.T., Olson, L.E., Tuljapurkar, S. & Coulson, T. (2010) Coupled dynamics of body mass and population growth in response to environmental change. *Nature* **466**, 482–485.

- 597 Peterson, M.L., Morris, W., Linares, C. & Doak, D. (2019) Improving structured population models with  
598 more realistic representations of non-normal growth. *Methods in Ecology and Evolution* **10**, 1431–1444.
- 599 Plard, F., Schindler, S., Arlettaz, R. & Schaub, M. (2018) Sex-specific heterogeneity in fixed morphological  
600 traits influences individual fitness in a monogamous bird population. *The American Naturalist* **191**, 106–119.
- 601 R Core Team (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical  
602 Computing, Vienna, Austria.
- 603 Rees, M., Childs, D.Z. & Ellner, S.P. (2014) Building integral projection models: a user's guide. *Journal  
of Animal Ecology* **83**, 528–545.
- 604
- 605 Rees, M. & Ellner, S.P. (2016) Evolving integral projection models: evolutionary demography meets  
606 eco-evolutionary dynamics. *Methods in Ecology and Evolution* **7**, 157–170.
- 607 Rigby, R.A., Stasinopoulos, M.D., Heller, G.Z. & De Bastiani, F. (2019) *Distributions for modeling location,  
scale, and shape: Using GAMLS in R*. CRC press.
- 608
- 609 Salguero-Gómez, R. & Casper, B.B. (2010) Keeping plant shrinkage in the demographic loop. *Journal  
of Ecology* **98**, 312–323.
- 610
- 611 Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D.F., Allegue, H., Teplitsky, C., Réale, D.,  
612 Dochtermann, N.A., Garamszegi, L.Z. & Araya-Ajoy, Y.G. (2020) Robustness of linear mixed-effects  
613 models to violations of distributional assumptions. *Methods in ecology and evolution* **11**, 1141–1152.
- 614 Shriver, R.K., Cutler, K. & Doak, D.F. (2012) Comparative demography of an epiphytic lichen: support  
615 for general life history patterns and solutions to common problems in demographic parameter estimation.  
616 *Oecologia* **170**, 137–146.

- 617 Stasinopoulos, D.M., Rigby, R.A. *et al.* (2007) Generalized additive models for location scale and shape  
618 (GAMLSS) in r. *Journal of Statistical Software* **23**, 1–46.
- 619 Stubberud, M.W., Vindenes, Y., Vøllestad, L.A., Winfield, I.J., Stenseth, N.C. & Langangen, Ø. (2019)  
620 Effects of size-and sex-selective harvesting: An integral projection model approach. *Ecology and  
621 Evolution* **9**, 12556–12570.
- 622 van der Laan, M.J. & Rose, S. (2011) *Targeted Learning: Causal Inference for Observational and  
623 Experimental Data*. Springer Science and Business Media.
- 624 Wan, X., Wang, W., Liu, J. & Tong, T. (2014) Estimating the sample mean and standard deviation from  
625 the sample size, median, range and/or interquartile range. *BMC medical research methodology* **14**, 1–13.
- 626 Williams, J.L., Miller, T.E. & Ellner, S.P. (2012) Avoiding unintentional eviction from integral projection  
627 models. *Ecology* **93**, 2008–2014.
- 628 Wood, S. (2017) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edn.
- 629 Zhang, J.L. (2014) Comparative investigation of three Bayesian *p* values. *Computational Statistics & Data  
630 Analysis* **79**, 277–291.

## Supporting Information

### 631 S.1 The Jones-Pewsey (2009) sinh-arcsinh distributions

632 Jones & Pewsey (2009) introduced a tractable generalization of the Normal distribution with two additional  
633 parameters determining asymmetry (skewness), and tail weight (kurtosis) which can be either lighter or  
634 heavier than the Gaussian. The generalization is defined through transformation of the Normal distribution  
635 using the hyperbolic sine function (sinh) and its inverse (asinh), as follows. The base distribution  $f_{\epsilon,\delta}$  is the  
636 probability density of the random variable  $X_{\epsilon,\delta}$  where

637 
$$Z = \sinh(\delta \operatorname{asinh}(X_{\epsilon,\delta}) - \epsilon) \quad (\text{S1})$$

638 and  $Z$  has a  $\mathcal{N}(0,1)$  distribution. Equivalently,

639 
$$X_{\epsilon,\delta} = \sinh(\delta^{-1} [\operatorname{asinh}(Z) + \epsilon]), \quad Z \sim \mathcal{N}(0,1). \quad (\text{S2})$$

640 Parameters  $\delta=1, \epsilon=0$  give the  $\mathcal{N}(0,1)$  distribution. Skewness has the sign of  $\epsilon$ , and  $\delta>0$  controls tail  
641 weight, with heavier than Gaussian tails for  $\delta<1$  and lighter than Gaussian tails for  $\delta>1$ . We show below  
642 that the nonparametric kurtosis (eqn. (3) in the main text) depends only on  $\delta$ , not on any of the other three  
643 parameters.

644 The density function for  $X_{\epsilon,\delta}$  is given by Jones & Pewsey (2009, eqn. 2),

645 
$$f_{\epsilon,\delta}(x) = C(x) \exp\{-S(x)^2/2\} \{2\pi(1+x^2)\}^{-1/2}$$

where  $S(x) = \sinh(\delta \operatorname{asinh}(x) - \epsilon)$ , (S3)

$$C(x) = \sqrt{1+S(x)^2} = \cosh(\delta \operatorname{asinh}(x) - \epsilon).$$

646 The attainable combinations of skewness and kurtosis are very broad compared to other families, and  
647 come very close to the theoretical limit of kurtosis as a function of skewness (Jones & Pewsey, 2009,  
648 Fig. 2). Additionally, eqn. (S2) makes it straightforward to generate random numbers and to compute  
649 the probability density, cumulative distribution, and quantile functions. There are also analytic formulas  
650 for the first four non-central moments (Jones & Pewsey, 2009, p. 764) in terms of the Bessel function  
651  $K_v$ , which is `BesselK` in base R and `besselk` in MATLAB and GNU OCTAVE.

652 Jones & Pewsey (2009) defined a four-parameter distribution with location parameter  $\mu$  and scale  
653 parameter  $\sigma$  as the distribution of  $\mu + \sigma X_{\epsilon,\delta}$ , which has density function

654 
$$f_{\mu,\sigma,\epsilon,\delta}(x) = \sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu)). \quad (\text{S4})$$

Terminology on this distribution has become somewhat confused. In the **mgcv** R package it is called `shash`, while in the **gamlss** package it is called `SHASHo2` and `SHASH` is a related but different distribution. To sidestep this confusion we refer to (S4) as the JP4 distribution (“Jones-Pewsey 4 parameter”), and refer to (S2) as JP2.

As is the case for most four-parameter distribution families, the location parameter  $\mu$  is not the mean of the JP4 distribution, and  $\sigma$  is not the standard deviation (additionally,  $\epsilon$  is not the skew and  $\delta$  is not the kurtosis). We therefore define a new four-parameter distribution family, JPS, by shifting and scaling JP2 so that the location parameter  $\mu$  is the mean, and the scale parameter  $\sigma$  is the standard deviation. This form can then be used in custom likelihood functions that “import” the fitted mean and standard deviation from a Gaussian pilot model, in the same way that the skewed  $t$  distribution was used in our lady orchid case study.

Let  $m(\epsilon, \delta)$  and  $s(\epsilon, \delta)$  denote the mean and standard deviation of the JP2 distribution. Then define

$$X_{JPS} = \mu + \sigma \left( \frac{X_{\epsilon, \delta} - m(\epsilon, \delta)}{s(\epsilon, \delta)} \right). \quad (\text{S5})$$

The right-hand term in parentheses has mean 0 and variance 1, so  $X_{JPS}$  has mean  $\mu$  and variance  $\sigma^2$ , with  $\epsilon$  and  $\delta$  controlling skewness and tail weight as in JP2. Because  $\mu$  and  $\sigma$  have no effect on the nonparametric kurtosis, JPR retains the property that nonparametric kurtosis only depends on  $\delta$ , not on the other three parameters.

Omitting some algebra,  $X_{JPS}$  has cumulative distribution function

$$\Pr(X_{JPS} \leq x) = \Pr\left(X_{\epsilon, \delta} \leq m(\epsilon, \delta) + \frac{s(\epsilon, \delta)}{\sigma}(x - \mu)\right). \quad (\text{S6})$$

Differentiating both sides with respect to  $x$ , the probability distribution function for  $X_{JPS}$  is

$$f_{JPS}(x|\mu, \sigma, \epsilon, \delta) = \frac{s(\epsilon, \delta)}{\sigma} f_{\epsilon, \delta}\left(m(\epsilon, \delta) + \frac{s(\epsilon, \delta)}{\sigma}(x - \mu)\right) \quad (\text{S7})$$

Eqn. (S2) shows that the JP2 distribution depends on  $\epsilon$  only through the ratio  $\epsilon/\delta$ , and hence the same is true for JPS. We have found that this property can be problematic for parameter estimation, because of the resulting ridge in the likelihood surface with constant  $\epsilon/\delta$ . Another problem is that when  $\delta$  is large, changes in  $\epsilon$  have little effect.

To avoid those problems, we recommend writing likelihood functions in terms of skewness and kurtosis parameters  $\lambda$  and  $\tau$ , defined by  $\delta = e^{-\tau}$ ,  $\epsilon = \delta\lambda$  in the JPS distribution. We will refer to this as the JPR distribution, with probability density

$$f_{JPR}(x|\mu, \sigma, \lambda, \tau) = f_{JPS}(x|\mu, \sigma, e^{-\tau}\lambda, e^{-\tau}). \quad (\text{S8})$$

682  $\lambda$  can take any real value, and the distribution's skewness has the same sign as  $\lambda$ .  $\tau$  also can take any  
 683 real value, with negative values giving thinner than Gaussian tails and positive values giving fatter than  
 684 Gaussian tails. Because  $\delta$  depends only on  $\tau$ , JPR also has the property that the nonparametric kurtosis  
 685 depends only on the tail-weight parameter  $\tau$ .

686 It is still the case that the ordinary skewness and kurtosis depend on both  $\lambda$  and  $\tau$ , but the “crosstalk”  
 687 is weaker than that between  $\epsilon$  and  $\delta$  (in particular, the tail-weight parameter has much less effect on the  
 688 skewness). As a result, we found that likelihood optimization is numerically more stable when the likelihood  
 689 function is written as a function of  $\tau$  and  $\lambda$  rather than  $\delta$  and  $\epsilon$ .

690 R code for the JP2, JPS, and JPR distributions with the usual `d,p,q,r` functions are provided in  
 691 the script `JP_funs.R` in our R code archive.

692 *Proof that NP kurtosis of JP2 depends only on  $\delta$ :* Let  $Z_\alpha$  denote the  $\alpha$  percentile of a standard Normal  
 693 distribution,  $X_\alpha$  the  $\alpha$  percentile of  $X_{\epsilon,\delta}$  and  $\lambda = \epsilon/\delta$ . Then from (S2) we have

$$694 \begin{aligned} X_{1-\alpha} &= \sinh[\lambda + \delta^{-1} \operatorname{asinh}(Z_{1-\alpha})], \\ X_\alpha &= \sinh[\lambda + \delta^{-1} \operatorname{asinh}(Z_\alpha)] = \sinh[\lambda + \delta^{-1} \operatorname{asinh}(-Z_{1-\alpha})] \\ &= \sinh[\lambda - \delta^{-1} \operatorname{asinh}(Z_{1-\alpha})]. \end{aligned} \quad (\text{S9})$$

695 Thus

$$696 X_{1-\alpha} - X_\alpha = \sinh(\lambda + b) - \sinh(\lambda - b) \quad (\text{S10})$$

697 where  $b = \delta^{-1} \operatorname{asinh}(Z_{1-\alpha})$ . We can apply the subtraction formula for sinh (eqn. 4.5.42 in Abramowitz  
 698 & Stegun (1970)), namely<sup>4</sup>

$$699 \sinh z_1 - \sinh z_2 = 2 \cosh\left(\frac{z_1 + z_2}{2}\right) \sinh\left(\frac{z_1 - z_2}{2}\right), \quad (\text{S11})$$

700 obtaining

$$701 X_{1-\alpha} - X_\alpha = 2 \cosh(\lambda) \sinh(b). \quad (\text{S12})$$

702 The value of  $b$  is independent of  $\epsilon$ . The  $\epsilon$ -dependent factor  $2 \cosh(\lambda)$  cancels in the numerator and  
 703 denominator of the formula for nonparametric kurtosis.  $\square$

## 704 S.2 Estimating random effects in non-Gaussian models using shrinkage

705 Specialized software for fitting mixed effects models only allow a subset, usually a small subset, of the  
 706 distributions that are useful for modeling growth.<sup>5</sup> One way past this limitation is Bayesian estimation.

---

<sup>4</sup>It's also on Wikipedia, and today it's correct, but tomorrow could be different.

<sup>5</sup>The `gamlss` package includes many distributions, but in our experience even with simple random effects structure the fitting algorithms often fail to converge reliably.

707 Here we describe another option, introduced by Link & Nichols (1994) and Gould & Nichols (1998):  
708 fitting the model in a fixed effects framework by Maximum Likelihood, followed by shrinkage of coefficient  
709 estimates. None of the ideas here are original. This section overlaps Appendix S1 of Metcalf *et al.* (2015),  
710 the only new wrinkle being the application to non-Gaussian models.

711 We explain shrinkage using a simple model fitted to some growth data on the bunchgrass *Pseudoroegneria*  
712 *spicata* from Adler *et al.* (2019). The fitted model includes random effects for across-year variation in the  
713 slope and intercept of future size (log area) as a function of initial size. We assume that initial size and year are  
714 the only covariates, and we assume that growth increments follow a skew-Normal distribution with nonconstant  
715 variance and constant skew parameter. Code for this example is in the script `SimpleShrinkageExample.R`.

716 The fitted growth model assumes that the skew and kurtosis parameters are functions of the location  
717 parameter; this dominated ( $\Delta AIC \approx 30$ ) the analogous model with skew and kurtosis depending on  
718 initial size. We fitted this model by MLE with all between-year variation appearing as fixed effects. The  
719 appropriate design matrix can be constructed using the `model.matrix` function:

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

721 If there are  $T$  years, the matrix `U` has  $2T$  columns corresponding to  $T$  annual intercepts and  $T$  annual slopes.

722 Using this design matrix, we can write a log-likelihood function for use with the `maxLik` package,  
723 using a log link function for the variance parameter because it is necessarily positive:

```
724 LogLik=function(pars,new.size,U){  
725   pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];  
726   mu = U%*%pars1;  
727   sigma = exp(pars2[1]+pars2[2]*mu);  
728   dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)  
729 }
```

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

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

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

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

$$740 \quad \hat{a}_t = a_t + \varepsilon_t \quad (\text{S13})$$

741 Because of sampling errors, the expected sample variance of the estimates  $\hat{a}_t$  is larger than the true  
 742 across-year variance in the parameter, which is undesirable if population projections are made by random  
 743 sampling from the estimated year-specific parameters (analogous to “matrix selection” for stochastic matrix  
 744 models). However, the approximate variance-covariance matrix  $\hat{V}$  of the sampling errors,  $V$  in the code  
 745 above, can be used to correct for this upward bias.

746 To make the correction we assume that the estimates  $\hat{a}_t$  are unbiased, that is

$$747 \quad \mathbb{E}(\varepsilon_t | a_t) = 0. \quad (\text{S14})$$

748 We also adopt the standard mixed-model assumption that the  $a_t$  are drawn independently from some  
 749 fixed distribution with unknown variance  $\sigma^2$ . These are optimistic assumptions, but not excessively so.  
 750 If the assumptions of maximum likelihood are satisfied, the bias in parameter estimates is asymptotically  
 751 negligible compared to the standard error. The terms resulting from non-independence can only be reliably  
 752 estimated if the autocorrelations fall to nearly zero within lag  $m \ll T$ , and in that case the autocorrelation  
 753 correction term is small (see eqn. (1) in Gould & Nichols (1998)). We therefore recommend proceeding  
 754 on the assumption that the  $\hat{a}_t$  are independent.

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

$$756 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (\text{S15})$$

757 This is equivalent to eqn. (1) in Gould & Nichols (1998) without the term that accounts for temporal  
 758 autocorrelation.

759 The terms besides  $\sigma^2$  on the right-hand of (S15) makes  $S^2$  a biased estimate of  $\sigma^2$ . However, those  
 760 terms correspond to entries in the variance-covariance matrix  $V$ , so we can use  $\hat{V}$  to remove the bias:

$$761 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (\text{S16})$$

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

763 We can similarly adjust the year-specific estimates to compensate for the expected impact of sampling  
 764 error. Several methods have been proposed; following Metcalf *et al.* (2015) we recommend the method

765 used in the capture-recapture analysis software Mark Cooch & White (2020, accessed 5/17/2020),

$$766 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (S17)$$

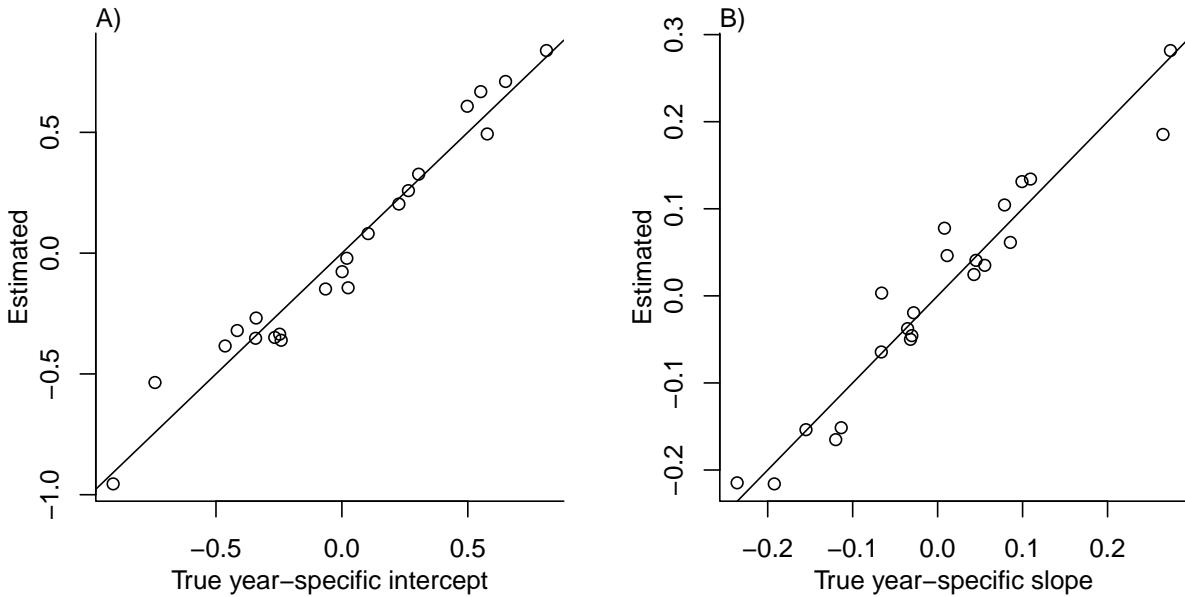
767 The name “shrinkage” comes from the fact that each estimate is adjusted towards the overall mean, with larger  
768 adjustments of values with higher estimated sampling error variance,  $\hat{V}_{t,t}$ . The expected sample variance of  
769 the adjusted estimates  $\tilde{a}_t$  is very close to  $\hat{\sigma}^2$ . The  $\tilde{a}_t$  therefore approximate the actual amount of parameter  
770 variation, and are analogous to the year-specific estimated random effects from a mixed effects model.

771 The take-home message is that estimating random effects from fitted year-specific regression  
772 coefficients is very simple. Continuing from the last code listing above:

```
773 # Variance-covariance matrices for intercepts and slopes
774 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
775 # Extract year-specific intercepts, center them to zero
776 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);

777
778 # Estimate sigma^2
779 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
780           (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
781
782 # Shrink deviations from the mean
783 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
784
785 # Do it all again for the slopes
786 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
787 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
788           (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
789 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));
```

790 Figure S-1 shows the results for one artificial “data” set, having  $T=22$  years and growth measurements  
791 on about 175 individuals per year on average. The true random year effects (that were used to generate  
792 the data) are recovered with good accuracy and no bias. In particular there is no sign of extreme values being  
793 pulled in too far towards the mean, which would cause an S-shaped graph of estimated versus true values.



**Figure S-1:** Comparison of the true random year effects with the shrinkage estimates, for one artificial data set generated from the fitted growth model for *Pseudoroegneria spicata*. Figure made by R script `SimpleShrinkageExample.R` in our code archive.

### 794 S.3 Additional case studies

#### 795 S.3.1 Sea fan corals, *Gorgonia ventalina*

796 Bruno *et al.* (2011) developed an IPM to understand the rise and fall of a fungal pathogen *Aspergillus sydowii*  
 797 in Caribbean sea fan corals *G. ventalina*. The model was based on repeated observations of marked corals in  
 798 permanent transects at several sites near Akumal, Mexico, recording disease status (infected/uninfected) and  
 799 the area of uninfected tissue. The epidemic peak had passed and disease incidence was already low, so  
 800 infected fans were relatively infrequent. We therefore limit the analysis here to uninfected individuals. Bruno  
 801 *et al.* (2011) found statistically significant year and site effects, but as those explained a very small fraction of  
 802 the variation in growth increments, they fitted a single growth model to data pooled across years and sites.  
 803 We do the same here. The pooled data set consists of 358 observed size transitions. The data exhibited  
 804 size-dependent variance in growth (change in area,  $cm^2$ ). Bruno *et al.* (2011) chose to stabilize the variance  
 805 by cube-root transforming size, and then fitting the standard model with Gaussian growth increments. Here  
 806 we take a different approach, using natural log transformation of area and modeling size-dependent variance.

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

```

811 # XH is a data frame holding the data
812 # logarea.t0, .t1 denote initial and final values of log-transformed area
813 fitGAU <- gam(list(logarea.t1~s(logarea.t0), ~ s(logarea.t0)),
814 data=XH, gamma=1.4, family=gaulss())
815 fitted_all = predict(fitGAU, type="response");
816 fitted_sd = 1/fitted_all[,2];
817 scaledResids = residuals(fitGAU, type='response')/fitted_sd;

```

818 Fig. S-3A shows the log-transformed data and Gaussian model. The mean function (solid red curve) is  
 819 visually nearly linear, but the fitted spline is strongly favored over a linear model for the mean ( $\Delta AIC \approx 9$ ).  
 820 The spline for standard deviation  $\sigma$  versus initial size reflects the evident greater variability in growth  
 821 at smaller sizes. Spline regression found only very small trends in the mean or variance of scaled residuals  
 822 (R script `crossssp_diagnose_pilot.R`; see Fig. S-2A,B).

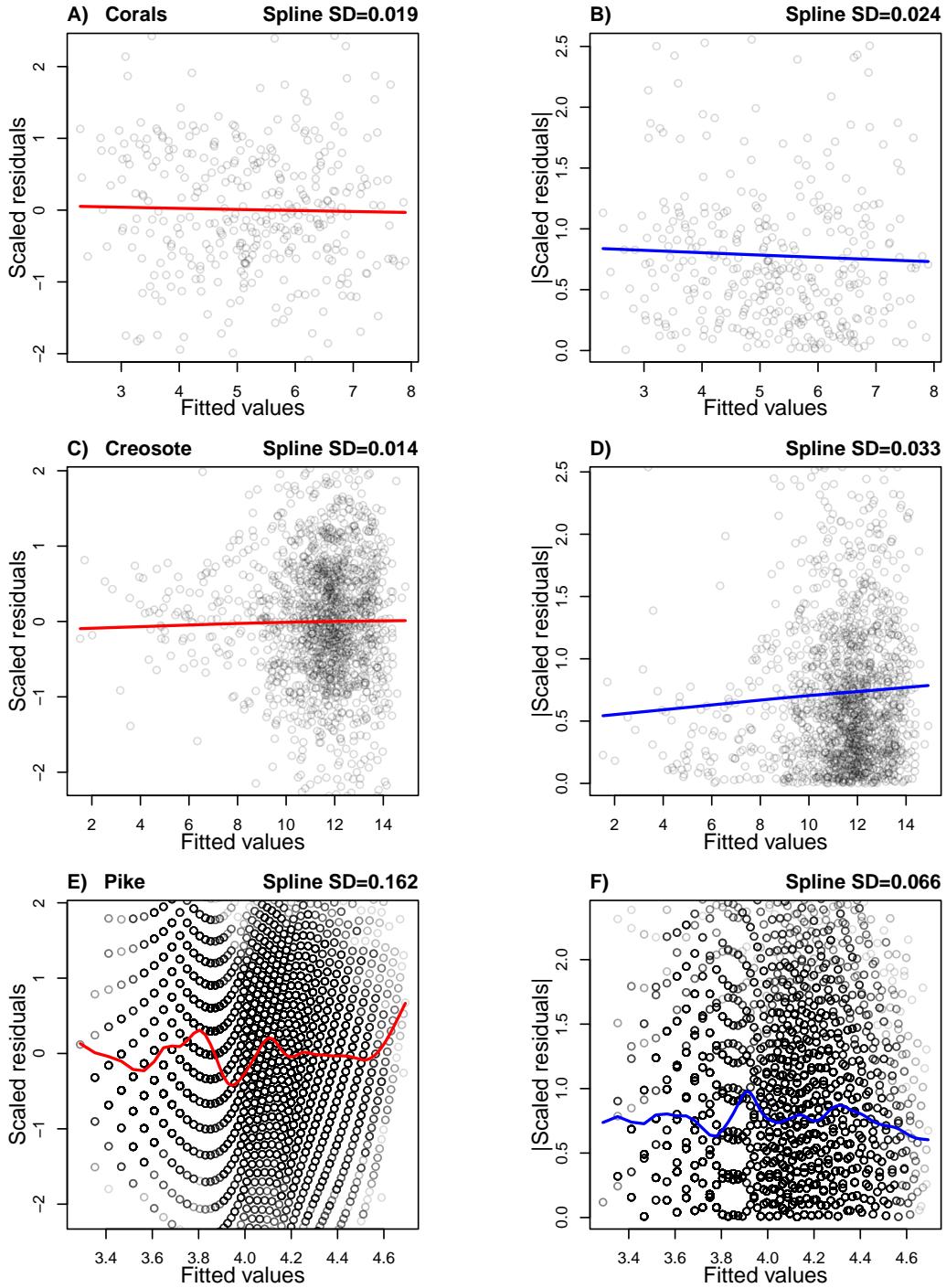
823 While there are no blatant signs of trouble in the pilot Gaussian model, quantile regressions on the scaled  
 824 residuals, and the NP Skewness and Kurtosis metrics derived from them (Eq. 2 and 3), suggest deviations from  
 825 normality (Fig. S-3B). Specifically, skewness switches from negative to positive across the size range, with  
 826 smaller corals more prone to extreme shrinkage and larger corals more prone to extreme growth. Kurtosis also  
 827 changes direction over the size distribution, with thinner tails than Gaussian at small sizes and fatter tails at large  
 828 sizes. The fitted nonparametric moments suggest that the upper and lower tails of size transition probabilities  
 829 may differ by up to 20%, and the weight of the tails may be 20% greater or less than Gaussian, depending  
 830 on initial size – not overwhelming deficiencies, but not trivial either. Are these deviations from normality  
 831 severe enough to warrant a second, non-Gaussian iteration of growth modeling? To answer that question, we  
 832 simulated data from the fitted Gaussian model and examined whether key properties of the simulated data are  
 833 consistent with those of the real data. If the simulated data are not consistent with the real data, it is time to  
 834 choose a better distribution (Fig. 1). In this case, most of 100 Gaussian model simulations are out of line with  
 835 the skew at smallest and largest sizes, and excess kurtosis observed at moderately large sizes (Fig. S-4 CD).  
 836 For at least some parts of the size distribution, a non-Gaussian model would better capture size transitions.

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

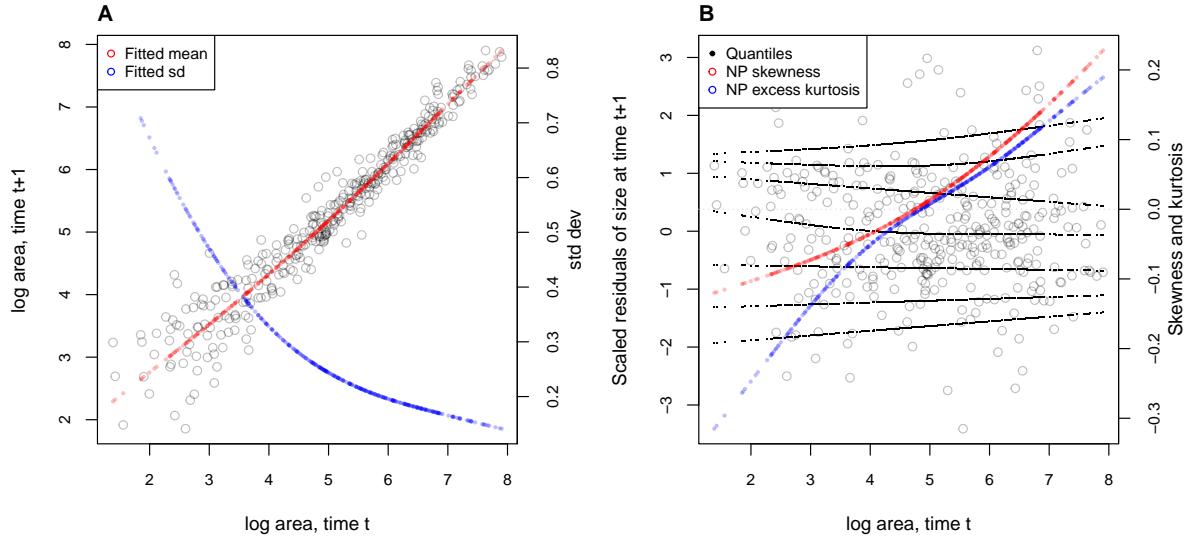
```

841 fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0), # <- location
842 ~ s(logarea.t0), # <- log-scale
843 ~ s(logarea.t0, k=4), # <- skewness

```



**Figure S-2:** Diagnostic plot for trends in the mean (left column) or variance (right column) of scaled residuals from a pilot Gaussian model, for the sea fan corals **A,B**, creosote bush **C,D**, and pike **E,F**. In **A,C,E** the standardized residuals are plotted, and in **B,D,F** the absolute values of standardized residuals, as functions of fitted mean subsequent size values. The solid curves are cubic splines (R function `smooth.spline`) fitted by generalized cross-validation with a modest over-penalization of model degrees of freedom to prevent overfitting (`penalty=1.4` as recommended by Gu (2013)). The numbers appearing above each panel are the standard deviation of the values on the spline regression curve, evaluated at all of the fitted values. Figure made by script `cross spp_diagnose_pilot.R`.



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

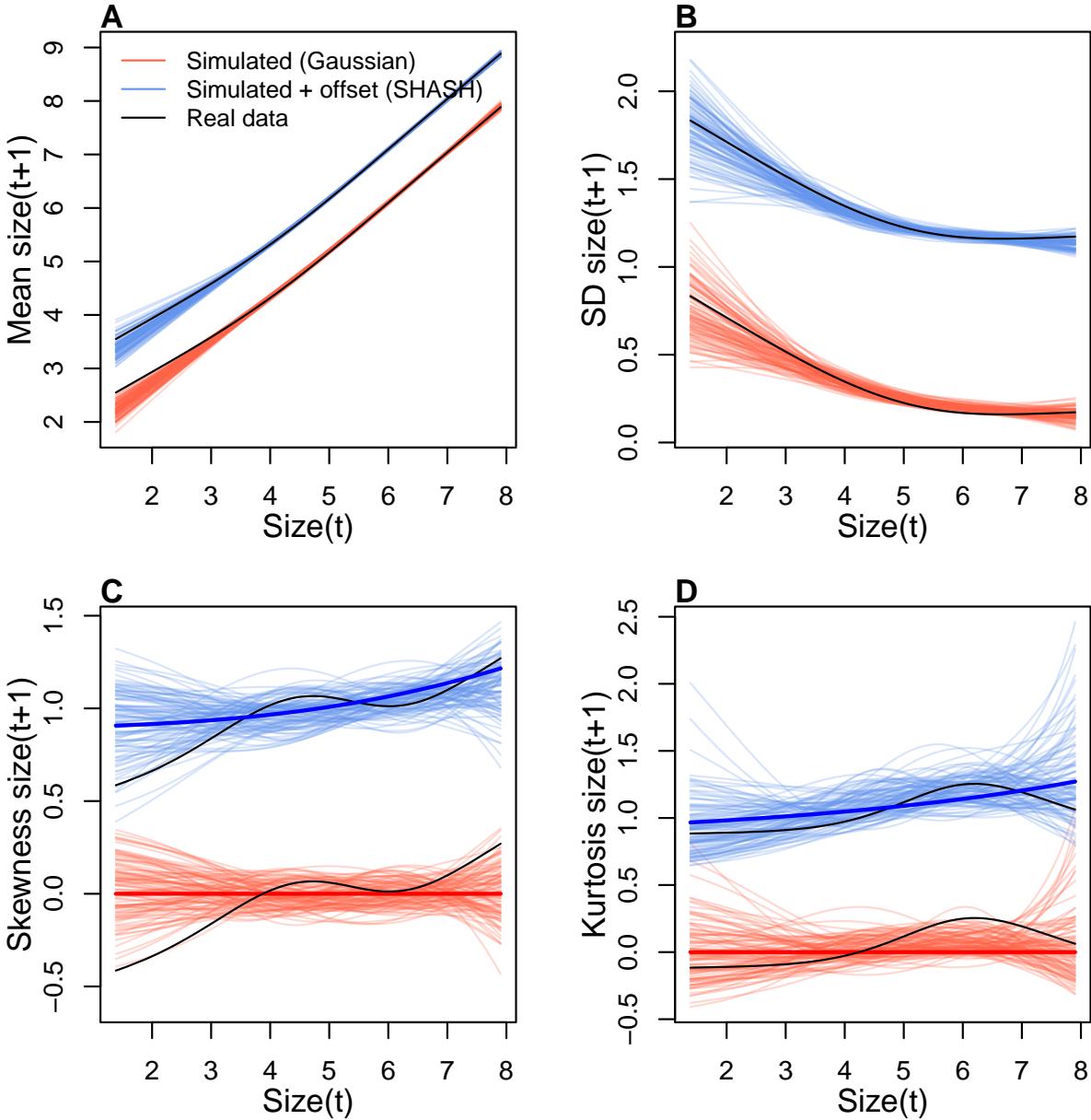
```

844 ~ s(logarea.t0,k=4)), # <- log-kurtosis
845 data = XH, gamma = 1.4, family = shash, optimizer = "efs")

```

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

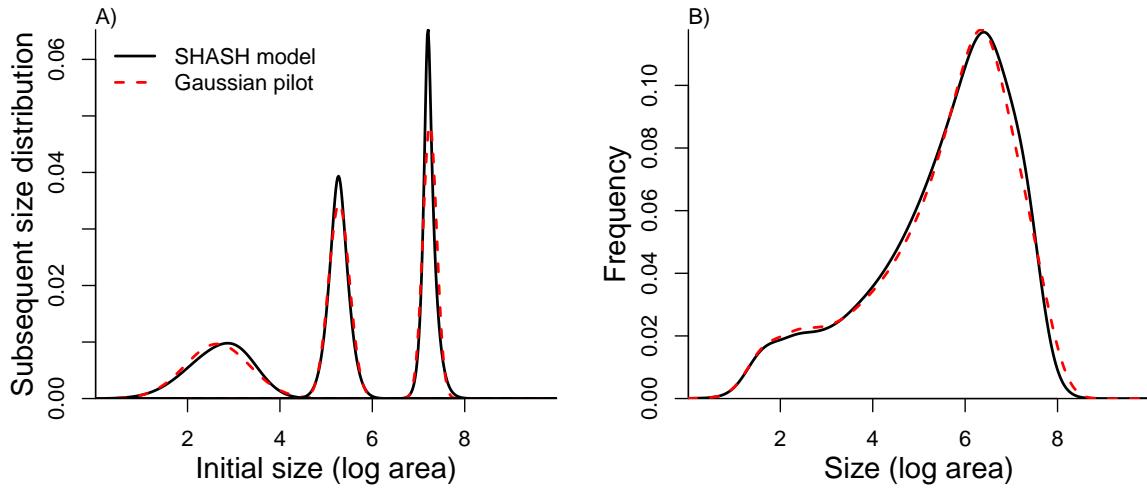
852 What, then, have we gained by fitting a better growth model? Fig. S-5A compares the predicted  
 853 distributions of subsequent size in the fitted model and Gaussian pilot models, for the median size of a new  
 854 recruit (leftmost pair of curves), the median initial size (central curves), and the 95th percentile of initial size  
 855 in the data (rightmost curves). The differences are small, and most pronounced for the smallest size, where  
 856 recruits are predicted to grow slightly larger under the SHASH model than the Gaussian model. The direction  
 857 of this difference was surprising, because the SHASH has negative skew at small sizes in the data. However,  
 858 the SHASH model also gives a better prediction of mean growth at small sizes than the Gaussian model. At  
 859 intermediate sizes the predictions are nearly identical; at large sizes the SHASH has slightly lower standard  
 860 deviation, but fatter tails (excess kurtosis). Fig. S-5B shows the predicted steady-state size distributions



**Figure S-4:** 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`.

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

In this case study we used `gam` to fit both the Gaussian and SHASH models because that obviated model selection on functions for mean, variance, and higher moments. However, `gam` should be used with caution. Nonparametric regression models notoriously “wag their tails” because the ends of the fitted



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

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

### 871 S.3.2 Creosotebush, *Larrea tridentata*

872 Our next case study comes from our studies of the woody shrub creosotebush (*Larrea tridentata*) at the  
 873 Sevilleta Long-Term Ecological Research (LTER) site in central New Mexico, US. At this site as elsewhere  
 874 in the Southwest US, creosotebush is encroaching into desert grassland habitats. The data described here  
 875 were collected along transects spanning grass-shrub ecotones to understand patterns of density dependence  
 876 in creosotebush demography. Specifically, we asked whether fitness is maximized approaching zero  
 877 density at the leading edge of the expansion front (consistent with ‘pulled’ expansion), or whether there  
 878 is a demographic advantage for shrubs at higher density due to positive feedbacks expected for ecosystem  
 879 engineers (leading to ‘pushed’ expansion). Our published study (Drees *et al.*, 2023) used a spatial integral  
 880 projection model (SIPM) to predict the speed of shrub encroachment, assuming normally-distributed size

881 transitions with non-constant variance. Here we ask whether a non-Gaussian model would have been more  
882 faithful to the data, and how such an improvement would influence predictions for the speed of encroachment.

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

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

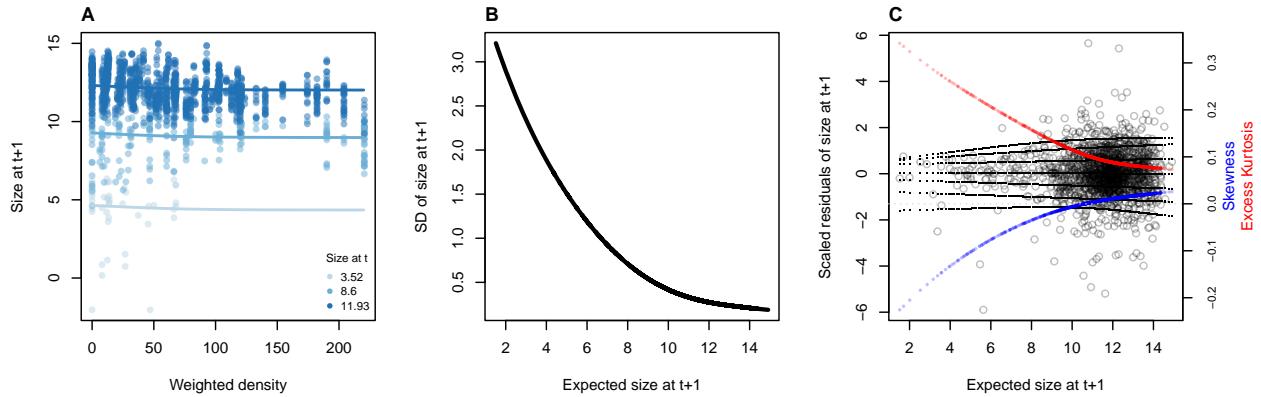
```
894 LATR_GAU <- gam(list(log_volume_t1~s(log_volume_t,k=4) +  
895 s(dens_scaled,k=4) + s(unique.transect,bs="re"), ~ 1),  
896 family="gaulss", data=LATR_grow, method="ML", gamma=1.4)
```

897 Using the fitted values from this initial model, we updated the standard deviation function to be a smooth  
898 function of fitted values, and iterated the fitting until the weights stopped changing, following the same steps  
899 as in the orchid case study. As with tree cholla cactus, the standard deviation function required  $k = 6$  basis  
900 functions to pass our graphical diagnostic (Fig. S-2C,D). The remaining small, nearly linear trend in the scale of  
901 standardized residuals (Fig. S-2D) is not improved by using  $k = 8$  basis functions, and appears to be driven by  
902 high leverage points in a region of relatively sparse data, so we did not attempt to further improve the pilot model.

903 The resulting Gaussian growth model predicts strong initial size-dependence and weak and slightly  
904 nonlinear (but monotonic) negative density dependence (Fig. S-6A). The model indicates non-constant  
905 variance, with greater dispersion at smaller sizes (Fig. S-6B).

906 Quantiles of the standardized residuals indicate that skew and excess kurtosis are both greater at  
907 smaller sizes (Fig. S-6C). Skewness is close to zero for larger plants (the best-sampled size range) but  
908 excess kurtosis remains positive for large plants (ca. 10% heavier tails than Gaussian). As a candidate  
909 for improvement, we turned to the Johnson's  $S_U$  (JSU) distribution, a four-parameter, leptokurtic distribution  
910 capable of skew in either direction.

911 Following our suggested modeling approach, rather than re-fitting a JSU model from scratch, we  
912 parameterize a model where the residuals from the Gaussian model are fitted by a JSU distribution. This  
913 is relatively easy because the **gamlss.dist** package provides a parameterization of the JSU in which the



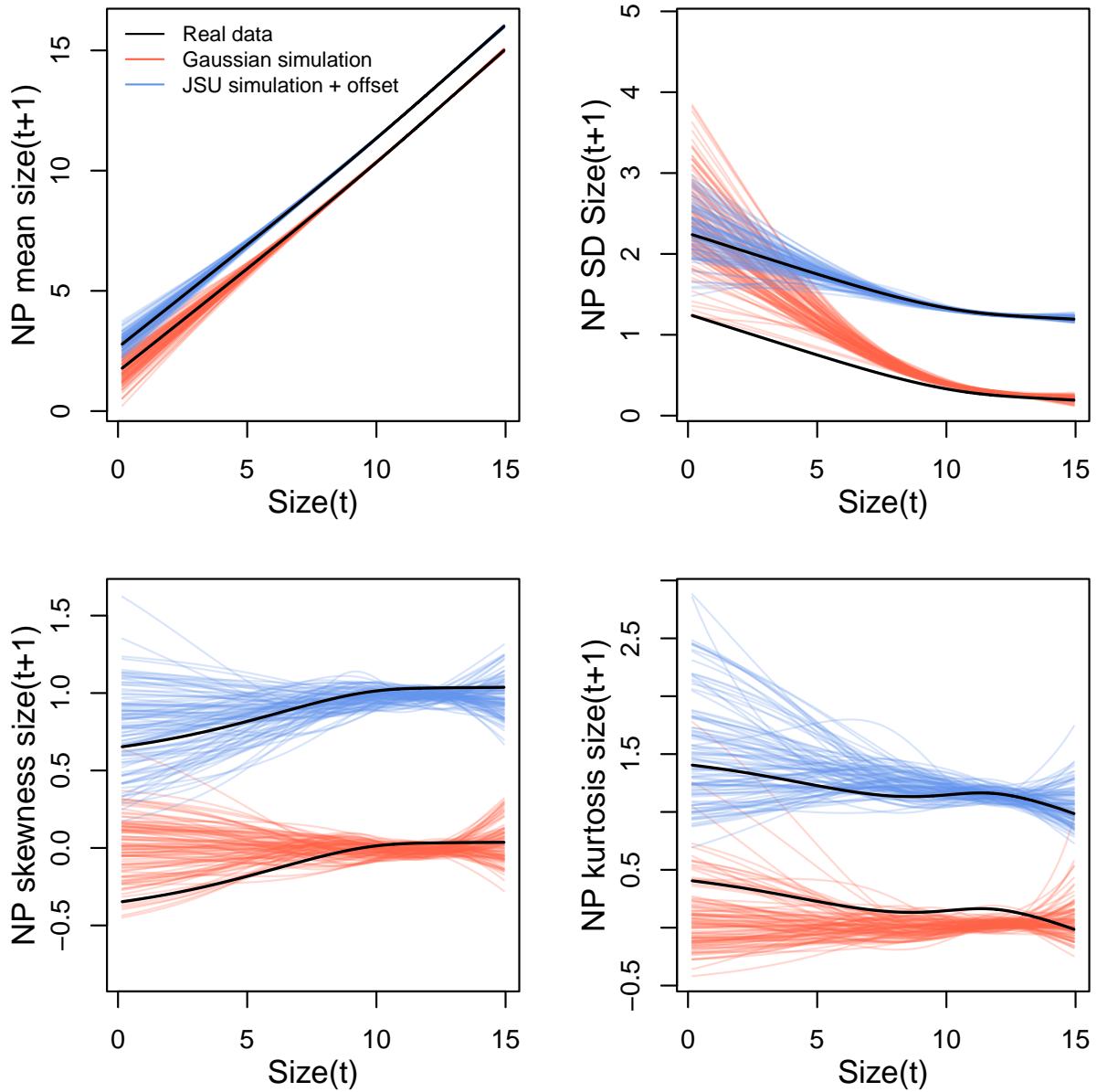
**Figure S-6:** **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume ( $\text{cm}^3$ ). **B**, Standard deviation of size 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`.

914 location parameter  $\mu$  is the mean and scale parameter  $\sigma$  is the standard deviation (Rigby *et al.*, 2019).  
 915 We fit the “hybrid” model by writing a likelihood function that uses the fitted mean and standard deviation  
 916 functions from Gaussian pilot model, and estimates the parameters that control skewness and kurtosis  
 917 as linear functions of predicted future size. The “hybrid” likelihood looks like this:

```
918 JSULogLik=function(pars){
  919   dJSU(LATR_grow$log_volume_t1,
  920         mu=LATR_grow$GAU_mean,
  921         sigma=LATR_grow$GAU_sd,
  922         nu = pars[1]+pars[2]*LATR_grow$GAU_mean,
  923         tau = exp(pars[3]+pars[4]*LATR_grow$GAU_mean), log=TRUE)
  924 }
```

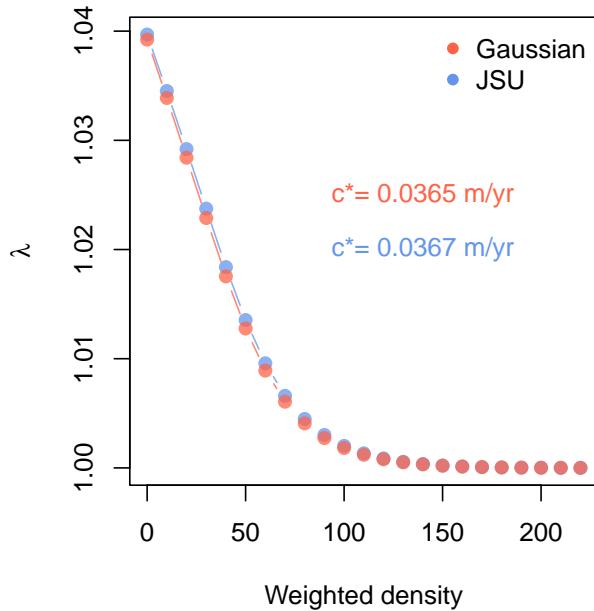
925 The mean and standard deviation of the JSU are set to those of the best Gaussian model and parameters  
 926 controlling skewness and kurtosis were fit independently, following our approach to the orchid data. The  
 927 hybrid JSU model performed well, generating simulated data that aligned with the real data better than the best  
 928 Gaussian model, particularly in the standard deviation and kurtosis (Fig. S-7). The JSU model has exactly  
 929 the same mean and standard deviation of future size as the Gaussian, but Fig. S-6 uses the quantile-based  
 930 nonparametric mean and standard deviation. The results show that even though the JSU was not fitted to  
 931 match those, it comes closer than the Gaussian model as a result of accounting for the skew and kurtosis.

932 The improvement of the JSU over the Gaussian growth model, while visually satisfying, had only weak  
 933 influence on SIPM results. The Gaussian model slightly over-estimated the low-density growth rate, but



**Figure S-7:** 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`.

models using either Gaussian or JSU growth kernels had very similar monotonic decreases in  $\lambda$  with increasing local density, and nearly identical wave velocities (Fig. S-8). This species has very low mortality risk once established (mean remaining life expectancy of a median-sized shrub is 24,408 years) and its population growth and wave expansion are limited by very low seedling recruitment ((Drees *et al.*, 2023)). Weak size-dependence in survival likely explains why the improvement in growth modeling had little influence on SIPM predictions.



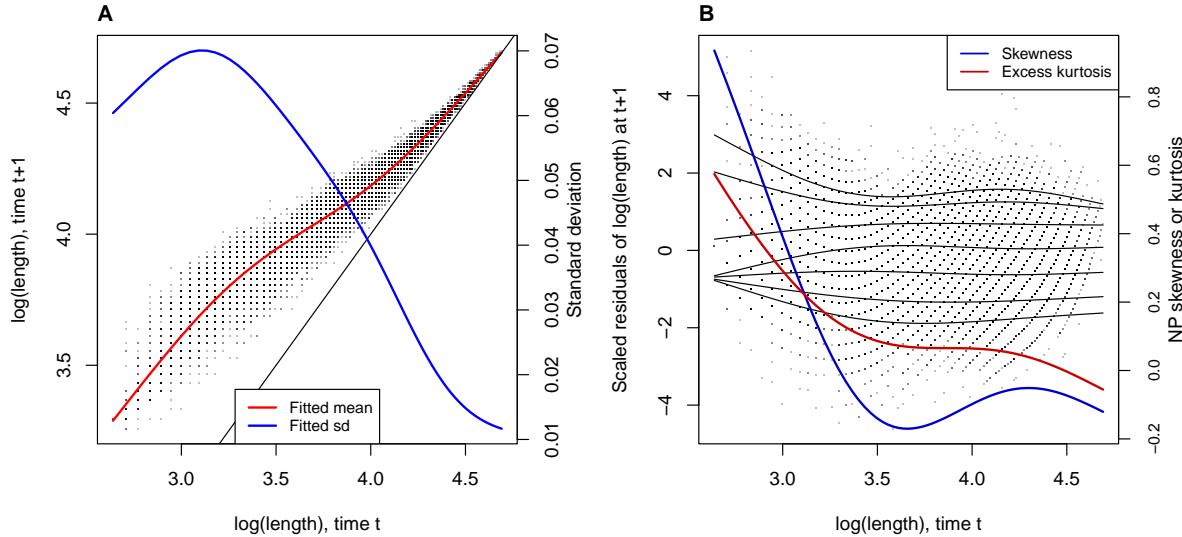
**Figure S-8:** Density dependence in fitness ( $\lambda$ ) and asymptotic velocity of the creosote encroachment wave ( $c^*$ ) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ( $\log(cm^3)$ ) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

### 939 S.3.3 Case study: pike, *Esox lucius*

940 Our final case study comes from a long-term (51 year) study of pike (*Esox lucius*) at Windemere in the English  
 941 Lake District, UK. Fish were gill-netted and destructively sampled to retrieve otoliths. Lengths (cm) were  
 942 recorded at the time of sampling and back-casted to estimate length in the preceding year. There were 26501 size  
 943 transitions in the data set. These data are publicly available (Winfield *et al.*, 2013b), as are data on size-specific  
 944 fertility and survival (Winfield *et al.*, 2013a,c), and have been analyzed in previous IPM studies (Vindenes *et al.*,  
 945 2014; Stubberud *et al.*, 2019). Previous authors modeled growth using a log-normal distribution to ensure  
 946 that change in length was non-negative. Here, we do not attempt to reproduce the published IPMs but rather  
 947 use the growth data as an additional test case of non-Gaussian growth modeling for a short-lived vertebrate.

948 With no additional covariates or random effects, this is a simple growth model of final size conditional  
 949 on initial size. We use the natural log of length. Our first step was a Gaussian model of  $\log(\text{length})$  where  
 950 the mean and standard deviation are smooth functions of initial size fit using the `gau1ss()` family in  
 951 `mgcv`. We then derive the scaled residuals from the fitted mean and standard deviation:

```
952 # pike is the data frame
953 #t1 and t0 are final and initial log(length), respectively
954 pike_gau<-gam(list(t1 ~ s(t0,k=5), ~ s(t0,k=5)), data=pike, family=gau1ss())
955 pike_gau_pred<-predict(pike_gau,type="response")
```



**Figure S-9:** **A**, Size transition data for pike, *Esox lucius*, 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.

```

956 pike$fitted_mean<-pike_gau_pred[,1]
957 pike$fitted_sd<-1/pike_gau_pred[,2]
958 pike$scaledResids=residuals(pike_gau,type="response")/fitted_sd

```

Based on preliminary fits we found that a basis function number of  $k = 5$  was necessary to minimize variance trends in the standardized residuals. Even so, because of the very large sample size, our graphical diagnostics for the pilot mean and standard deviation functions (Fig. S-2E,F) detected small-scale deviations from constant mean and variance. Note that individual sizes in this study were recorded somewhat coarsely (nearest 1cm). That accounts for the striking visual patterns in the scaled residuals, and probably also accounts for the small-scale patterns in the diagnostic regression curves (Fig. S-2E,F). In order to remove the small-scale wiggles in the diagnostic splines, we would have to introduce small-scale wiggles in the mean and variance functions, which are unlikely to be real features of pike growth trajectories. So while we are not entirely satisfied with our pilot model, we see no way to improve it.

The estimate growth variance strongly decreased with initial size, and size transitions were strongly positively skewed, with up to a 75% difference in tail weight at small sizes (Fig. S-9B). Size transitions were fat-tailed at small initial sizes but were consistent with Gaussian tails at large initial sizes.

Our improved growth model was a SHASH gam that defined all four parameters as smooth functions of initial size.

```

973 pike_gam_shash <- gam(list(t1 ~ s(t0,k=5), # <- model for location

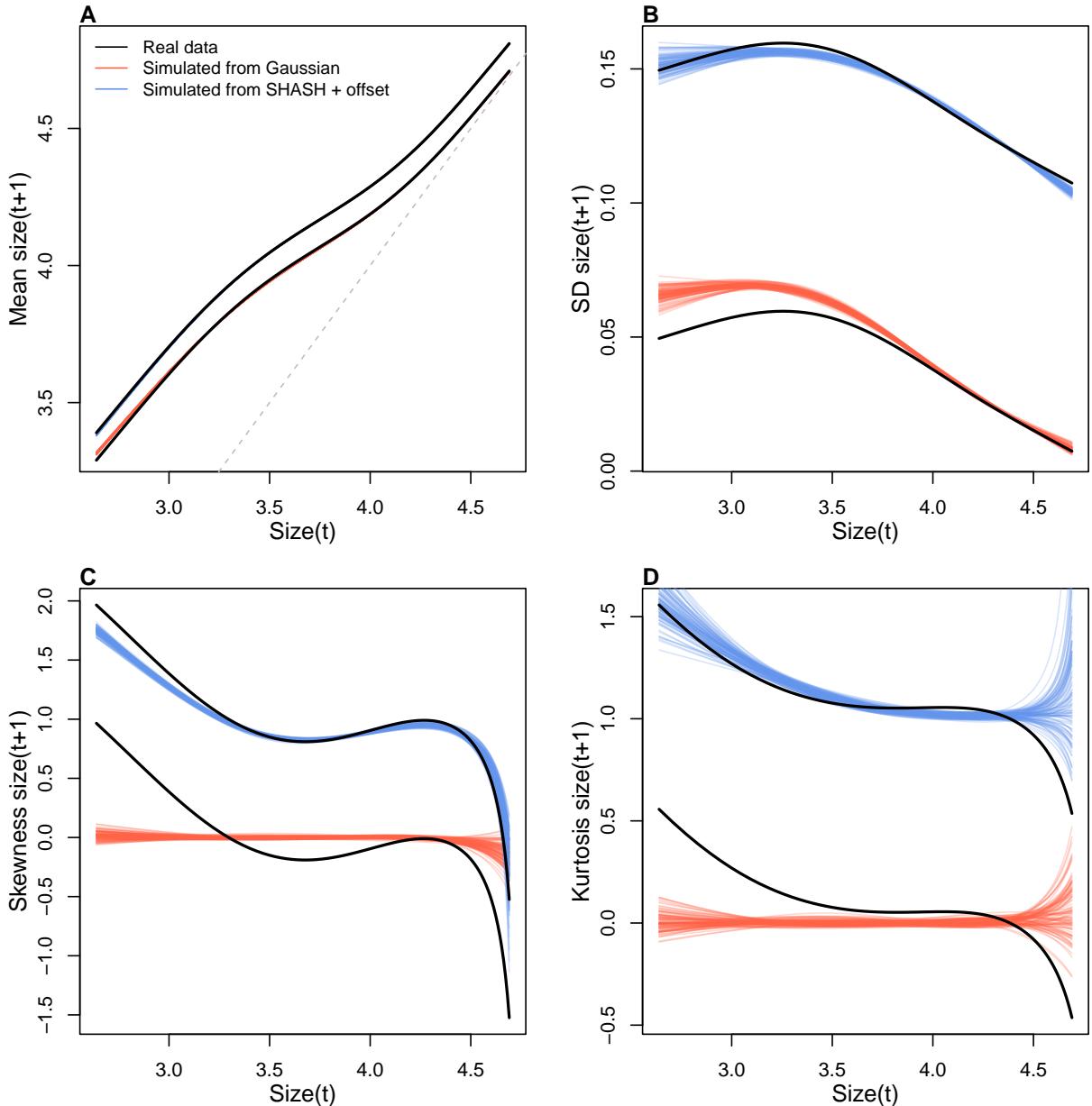
```

```
974 ~ s(t0,k=5), # <- model for log-scale  
975 ~ s(t0,k=5), # <- model for skewness  
976 ~ s(t0,k=5)), # <- model for log-kurtosis  
977 data = pike, family = shash, optimizer = "efs")
```

978 We also tried gamma regression on the change in size, to ensure strictly increasing size transitions, but found  
979 that this was not actually necessary to prevent shrinkage and did not provide as good a fit as the SHASH. Data  
980 simulated from the SHASH and Gaussian models are shown in Fig. S-10. The SHASH is an improvement  
981 over the Gaussian for most initial sizes. It fails to capture kurtosis of the largest fish, but that will have little  
982 effect because the fitted mean and standard deviation imply, correctly, that those fish will have very small and  
983 nearly deterministic growth increments until they reach the size at which growth ceases (Figs. S-9A, S-10A).

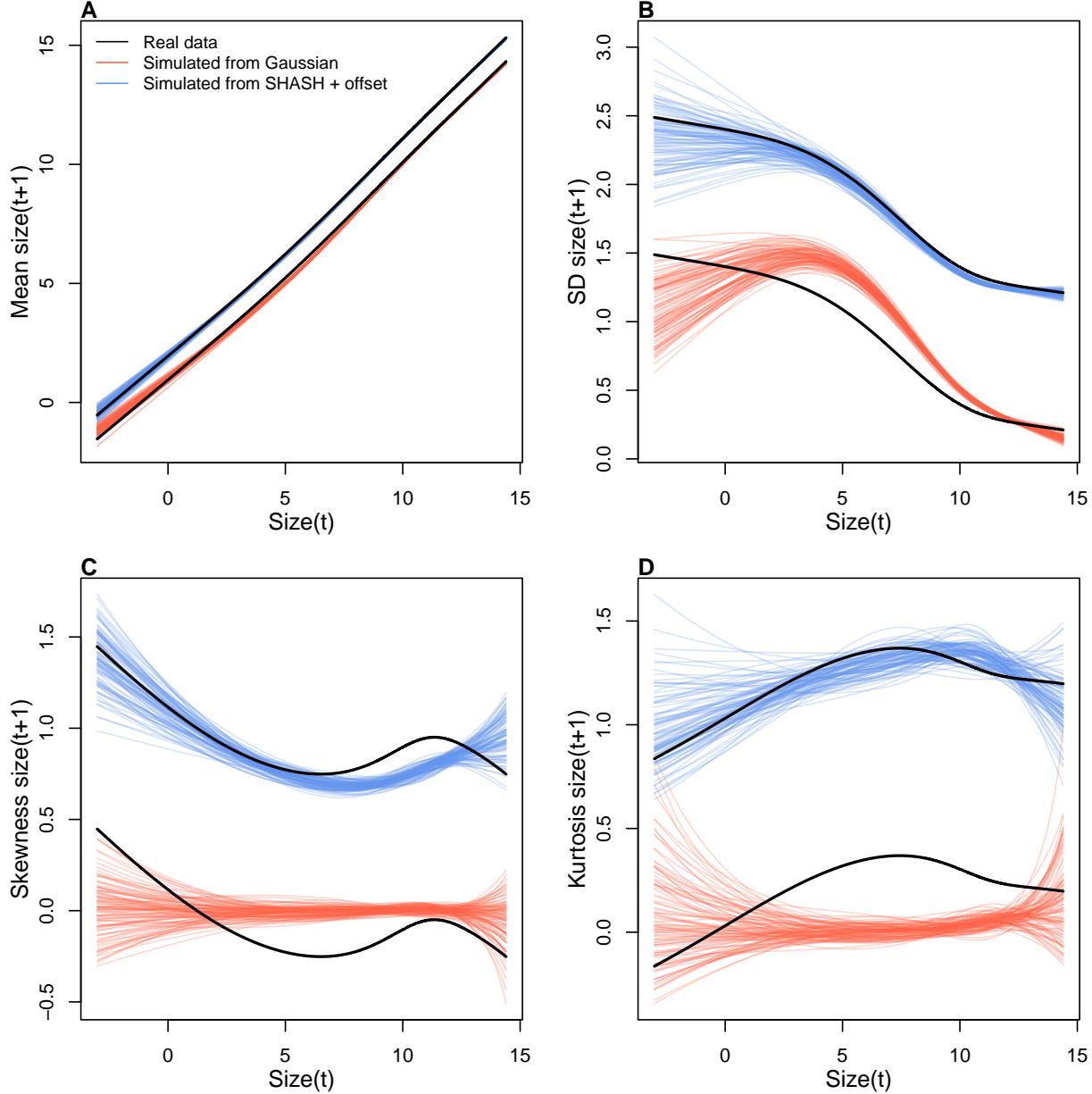
984 For the other components of the IPM, we fit GAMs for survival and egg production as smooth  
985 functions of size. Parameter values for fertilization probability, fraction female (the IPM is female-dominant),  
986 and probability of survival from egg to 1-yo came from Stubberud *et al.* (2019), Table 2.

987 Predictions from the SHASH- and Gaussian-growth IPMs (Table 1) are uniformly remarkably similar.

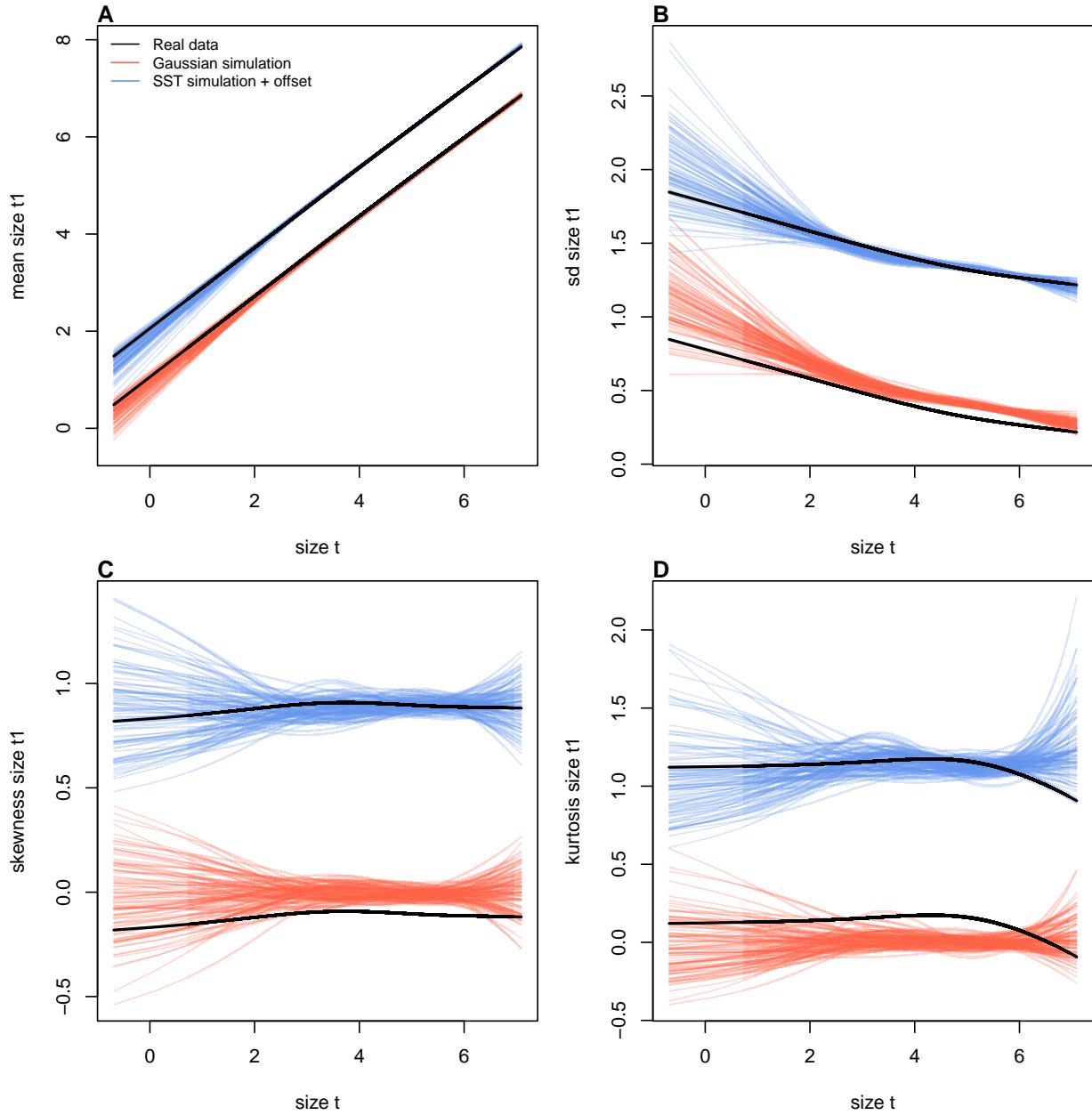


**Figure S-10:** Comparisons between real pike data and data simulated from Gaussian and SHASH growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect to initial size. The dashed line in the top-left panel is the 1:1 line. Figure made by script `PikeGrowthModeling_qgam.R`.

988 S.4 Additional Figures



**Figure S-11:** 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`.



**Figure S-12:** 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`.

989 **Literature Cited**

- 990 Abramowitz, M. & Stegun, I.A. (1970) Handbook of Mathematical Functions. 9th printing. Dover  
991 Publications, Inc., New York.
- 992 Bruno, J.F., Ellner, S.P., Vu, I., Kim, K. & Harvell, C.D. (2011) Impacts of aspergillosis on sea fan coral  
993 demography: modeling a moving target. *Ecological Monographs* **81**, 123–139.
- 994 Drees, T., Ochocki, B.M., Collins, S.L. & Miller, T.E. (2023) Demography and dispersal at a grass-shrub eco-  
995 tone: a spatial integral projection model for woody plant encroachment. *Ecological Monographs* p. e1574.
- 996 Gould, W.R. & Nichols, J.D. (1998) Estimation of temporal variability of survival in animal populations.  
997 *Ecology* **79**, 2531 – 2538.
- 998 Jones, M. & Pewsey, A. (2009) Sinh-arcsinh distributions. *Biometrika* **96**, 761 – 780.
- 999 Jones, M.C., Rosco, J.F. & Pewsey, A. (2011) Skewness-invariant measures of kurtosis. *The American*  
1000 *Statistician* **65**, 89 – 95.
- 1001 Link, W.A. & Nichols, J.D. (1994) On the importance of sampling variance to investigations of temporal  
1002 variation in animal population size. *Oikos* **69**, 539 – 544.
- 1003 Ochocki, B.M., Drees, T. & Miller, T.E. (2023) Density-dependent demography of creosote  
1004 bush (*larrea tridentata*) along grass-shrub ecotones. [https://doi.org/10.6073/pasta/  
1005 ca53c16f16dcf9fb11f3ee99ea5445ac](https://doi.org/10.6073/pasta/ca53c16f16dcf9fb11f3ee99ea5445ac).
- 1006 Vindenes, Y., Edeline, E., Ohlberger, J., Langangen, Ø., Winfield, I.J., Stenseth, N.C. & Vøllestad, L.A.  
1007 (2014) Effects of climate change on trait-based dynamics of a top predator in freshwater ecosystems.  
1008 *The American Naturalist* **183**, 243–256.
- 1009 Winfield, I., Fletcher, J. & James, J. (2013a) Pike fecundity data 1963-2002. NERC Environmental Information  
1010 Data Centre, <https://doi.org/10.5285/b8886915-14cb-44df-86fa-7ab718acf49a>.
- 1011 Winfield, I., Fletcher, J. & James, J. (2013b) Pike growth data 1944-1995. NERC Environmental Information  
1012 Data Centre, <https://doi.org/10.5285/637d60d6-1571-49af-93f7-24c1279d884d>.
- 1013 Winfield, I., Fletcher, J. & James, J. (2013c) Pike survival data 1953-1990. NERC Environmental Information  
1014 Data Centre, <https://doi.org/10.5285/813e07dd-2135-49bc-93c6-83999e442b36>.