

Appendix S1

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

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

$$7 \quad Z = \sinh(\delta \operatorname{asinh}(X_{\epsilon,\delta}) - \epsilon) \quad (\text{S1})$$

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

$$9 \quad X_{\epsilon,\delta} = \sinh(\delta^{-1} [\operatorname{asinh}(Z) + \epsilon]), \quad Z \sim \mathcal{N}(0,1). \quad (\text{S2})$$

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

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

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

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

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

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

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

$$24 \quad f_{\mu,\sigma,\epsilon,\delta}(x) = \sigma^{-1} f_{\epsilon,\delta}(\sigma^{-1}(x-\mu)). \quad (\text{S4})$$

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

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

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

$$36 X_{JPS} = \mu + \sigma \left(\frac{X_{\epsilon, \delta} - m(\epsilon, \delta)}{s(\epsilon, \delta)} \right). \quad (S5)$$

37 The right-hand term in parentheses has mean 0 and variance 1, so X_{JPS} has mean μ and variance σ^2 , with ϵ and
 38 δ controlling skewness and tail weight as in JP2. Because μ and σ have no effect on the nonparametric kurtosis,
 39 JPR retains the property that nonparametric kurtosis only depends on δ , not on the other three parameters.

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

$$41 Pr(X_{JPS} \leq x) = Pr\left(X_{\epsilon, \delta} \leq m(\epsilon, \delta) + \frac{s(\epsilon, \delta)}{\sigma}(x - \mu)\right). \quad (S6)$$

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

$$43 f_{JPS}(x | \mu, \sigma, \epsilon, \delta) = \frac{s(\epsilon, \delta)}{\sigma} f_{\epsilon, \delta}\left(m(\epsilon, \delta) + \frac{s(\epsilon, \delta)}{\sigma}(x - \mu)\right) \quad (S7)$$

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

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

$$51 f_{JPR}(x | \mu, \sigma, \lambda, \tau) = f_{JPS}(x | \mu, \sigma, e^{-\tau} \lambda, e^{-\tau}). \quad (S8)$$

52 λ can take any real value, and the distribution's skewness has the same sign as λ . τ also can take any
 53 real value, with negative values giving thinner than Gaussian tails and positive values giving fatter than
 54 Gaussian tails. Because δ depends only on τ , JPR also has the property that the nonparametric kurtosis
 55 depends only on the tail-weight parameter τ .

56 It is still the case that the ordinary skewness and kurtosis depend on both λ and τ , but the “crosstalk”
 57 is weaker than that between ϵ and δ (in particular, the tail-weight parameter has much less effect on the
 58 skewness). As a result, we found that likelihood optimization is numerically more stable when the likelihood
 59 function is written as a function of τ and λ rather than δ and ϵ .

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

62 *Proof that NP kurtosis of JP2 depends only on δ :* Let Z_α denote the α percentile of a standard Normal
 63 distribution, X_α the α percentile of $X_{\epsilon,\delta}$ and $\lambda = \epsilon/\delta$. Then from (S2) we have

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

65 Thus

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

66 where $b = \delta^{-1} \operatorname{asinh}(Z_{1-\alpha})$. We can apply the subtraction formula for sinh (eqn. 4.5.42 in Abramowitz
 67 & Stegun (1970)), namely¹

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

70 obtaining

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

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

¹It's also on Wikipedia, and today it's correct, but tomorrow could be different.

Appendix S2

74 S.1 Estimating random effects in non-Gaussian models using shrinkage

75 Specialized software for fitting mixed effects models only allow a subset, usually a small subset, of the
76 distributions that are useful for modeling growth.² One way past this limitation is Bayesian estimation.
77 Here we describe another option, introduced by Link & Nichols (1994) and Gould & Nichols (1998):
78 fitting the model in a fixed effects framework by Maximum Likelihood, followed by shrinkage of coefficient
79 estimates. None of the ideas here are original. This section overlaps Appendix S1 of Metcalf *et al.* (2015),
80 the only new wrinkle being the application to non-Gaussian models.

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

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

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

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

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

```
94 LogLik=function(pars,new.size,U){  
95     pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];  
96     mu = U%*%pars1;  
97     sigma = exp(pars2[1]+pars2[2]*mu);  
98     dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)  
99 }
```

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

```
101 start=c(runif(ncol(U)), rep(0,3))
```

²The `gamlss` package includes many distributions, but in our experience even with simple random effects structure the fitting algorithms often fail to converge reliably.

```

102     out=maxLik(logLik=LogLik,start=start, new.size=simData$new.size,U=U,
103     method="BHHH",control=list(iterlim=5000,printLevel=1),finalHessian=TRUE);
104     coefs = out$estimate; # parameters
105     V = vcov(out); SEs = sqrt(diag(V)); # standard errors

```

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

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

$$110 \quad \hat{a}_t = a_t + \varepsilon_t \quad (S1)$$

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

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

$$117 \quad \mathbb{E}(\varepsilon_t | a_t) = 0. \quad (S2)$$

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

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

$$126 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (S3)$$

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

129 The terms besides σ^2 on the right-hand of (S3) makes S^2 a biased estimated of σ^2 . However, those
 130 terms correspond to entries in the variance-covariance matrix V , so we can use \hat{V} to remove the bias:

$$131 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (S4)$$

132 $\hat{\sigma}^2$ is the estimated variance of the distribution from which the a_t are assumed to be drawn.

133 We can similarly adjust the year-specific estimates to compensate for the expected impact of sampling
 134 error. Several methods have been proposed; following Metcalf *et al.* (2015) we recommend the method
 135 used in the capture-recapture analysis software Mark Cooch & White (2020, accessed 5/17/2020),

$$136 \quad \tilde{a}_t = \bar{a}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{a}_t). \quad (S5)$$

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

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

```
143 # Variance-covariance matrices for intercepts and slopes
144 V1 = V[1:T,1:T]; V2 = V[(T+1):(2*T),(T+1):(2*T)];
145 # Extract year-specific intercepts, center them to zero
146 fixed.fx = coefs[1:T]; fixed.fx = fixed.fx-mean(fixed.fx);
147
148 # Estimate sigma^2
149 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
150 (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
151
152 # Shrink deviations from the mean
153 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
154
155 # Do it all again for the slopes
156 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
157 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
```

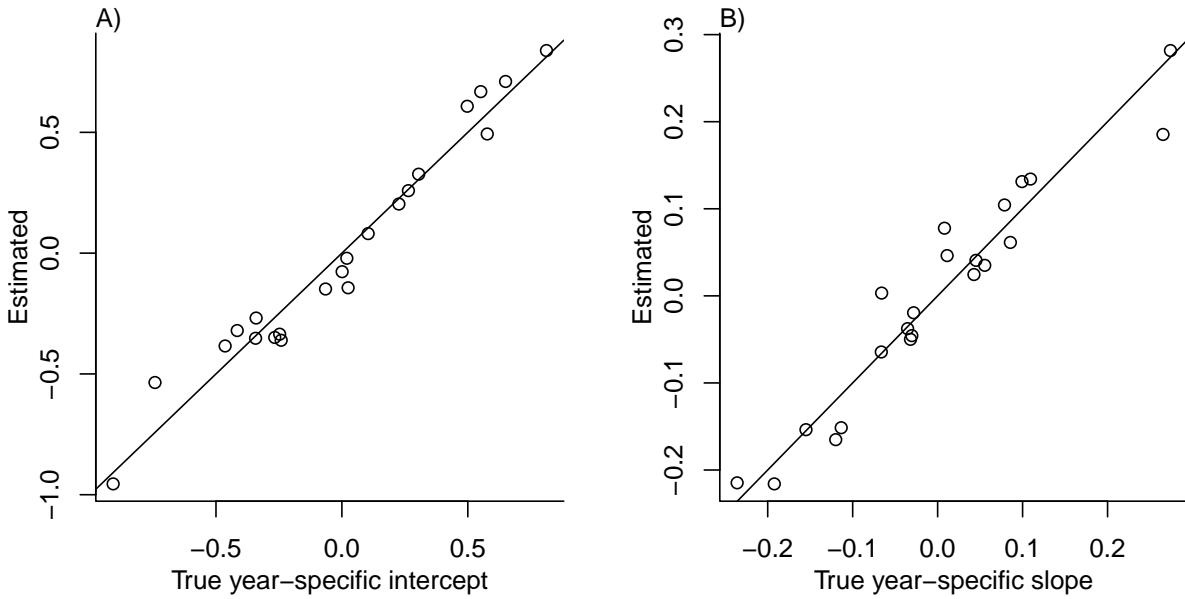


Figure S-1: Comparison of the true random year effects with the shrinkage estimates, for one artificial data set generated from the fitted growth model for *Pseudoroegneria spicata*. Figure made by R script `SimpleShrinkageExample.R` in our code archive.

```

158   (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
159   shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

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

Appendix S3

164 S.1 Additional case studies

165 S.1.1 Sea fan corals, *Gorgonia ventalina*

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

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

```
181 # XH is a data frame holding the data
182 # logarea.t0, .t1 denote initial and final values of log-transformed area
183 fitGAU <- gam(list(logarea.t1~s(logarea.t0), ~ s(logarea.t0)),
184   data=XH, gamma=1.4, family=gaulss())
185 fitted_all = predict(fitGAU, type="response");
186 fitted_sd = 1/fitted_all[,2];
187 scaledResids = residuals(fitGAU, type='response')/fitted_sd;
```

188 Fig. S-2A shows the log-transformed data and Gaussian model. The mean function (solid red curve) is
189 visually nearly linear, but the fitted spline is strongly favored over a linear model for the mean ($\Delta AIC \approx 9$).
190 The spline for standard deviation σ versus initial size reflects the evident greater variability in growth
191 at smaller sizes. Spline regression found only very small trends in the mean or variance of scaled residuals
192 (R script `crossssp_diagnose_pilot.R`; see Fig. S-1A,B).

193 While there are no blatant signs of trouble in the pilot Gaussian model, quantile regressions on the scaled
194 residuals, and the NP Skewness and Kurtosis metrics derived from them (Eq. 1 and 3 of the main text), suggest

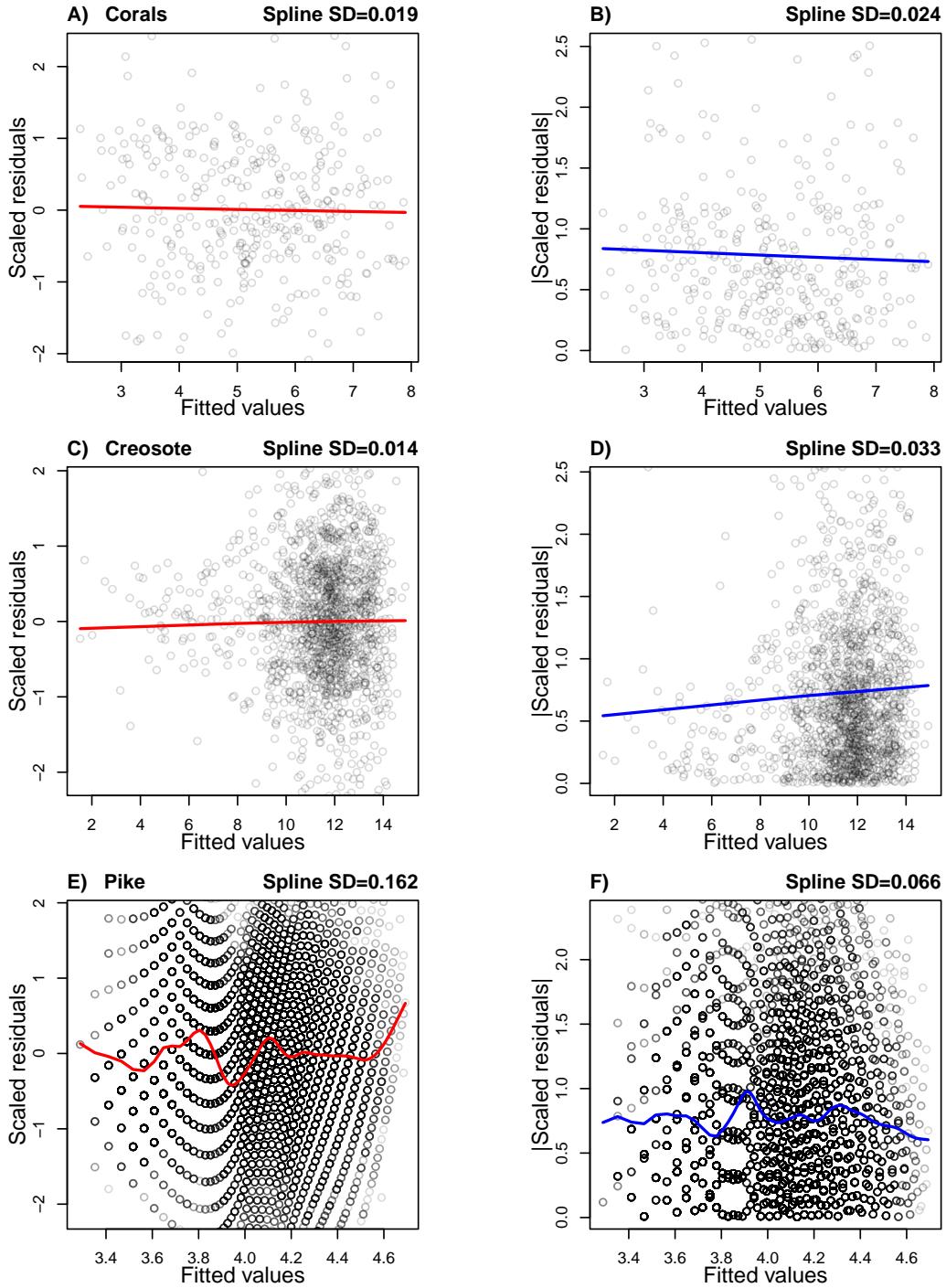


Figure S-1: Diagnostic plot for trends in the mean (left column) or variance (right column) of scaled residuals from a pilot Gaussian model, for the sea fan corals **A,B**, creosote bush **C,D**, and pike **E,F**. In **A,C,E** the standardized residuals are plotted, and in **B,D,F** the absolute values 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`.

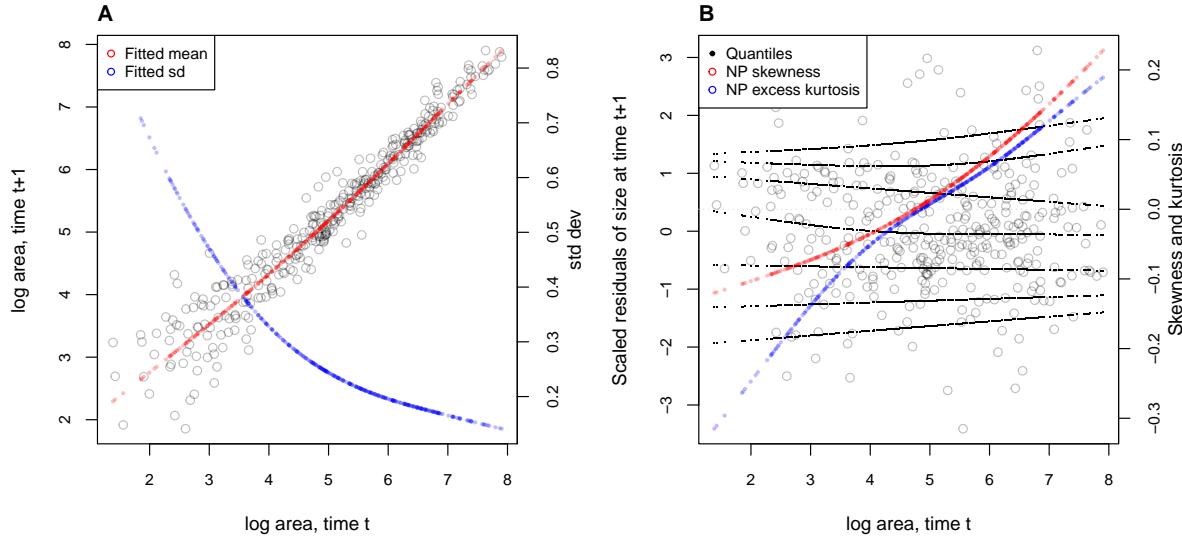


Figure S-2: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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`.

195 deviations from normality (Fig. S-2B). Specifically, skewness switches from negative to positive across the
 196 size range, with smaller corals more prone to extreme shrinkage and larger corals more prone to extreme growth.
 197 Kurtosis also changes direction over the size distribution, with thinner tails than Gaussian at small sizes and fatter tails at large sizes. The fitted nonparametric moments suggest that the upper and lower tails of size transition
 199 probabilities may differ by up to 20%, and the weight of the tails may be 20% greater or less than Gaussian,
 200 depending on initial size – not overwhelming deficiencies, but not trivial either. Are these deviations from nor-
 201 mality severe enough to warrant a second, non-Gaussian iteration of growth modeling? To answer that question,
 202 we simulated data from the fitted Gaussian model and examined whether key properties of the simulated data
 203 are consistent with those of the real data. If the simulated data are not consistent with the real data, it is time to
 204 choose a better distribution (Fig. 1). In this case, most of 100 Gaussian model simulations are out of line with
 205 the skew at smallest and largest sizes, and excess kurtosis observed at moderately large sizes (Fig. S-3 CD).
 206 For at least some parts of the size distribution, a non-Gaussian model would better capture size transitions.

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

```
211 fitSHASH <- gam(list(logarea.t1 ~ s(logarea.t0), # <- location  

  212 ~ s(logarea.t0), # <- log-scale
```

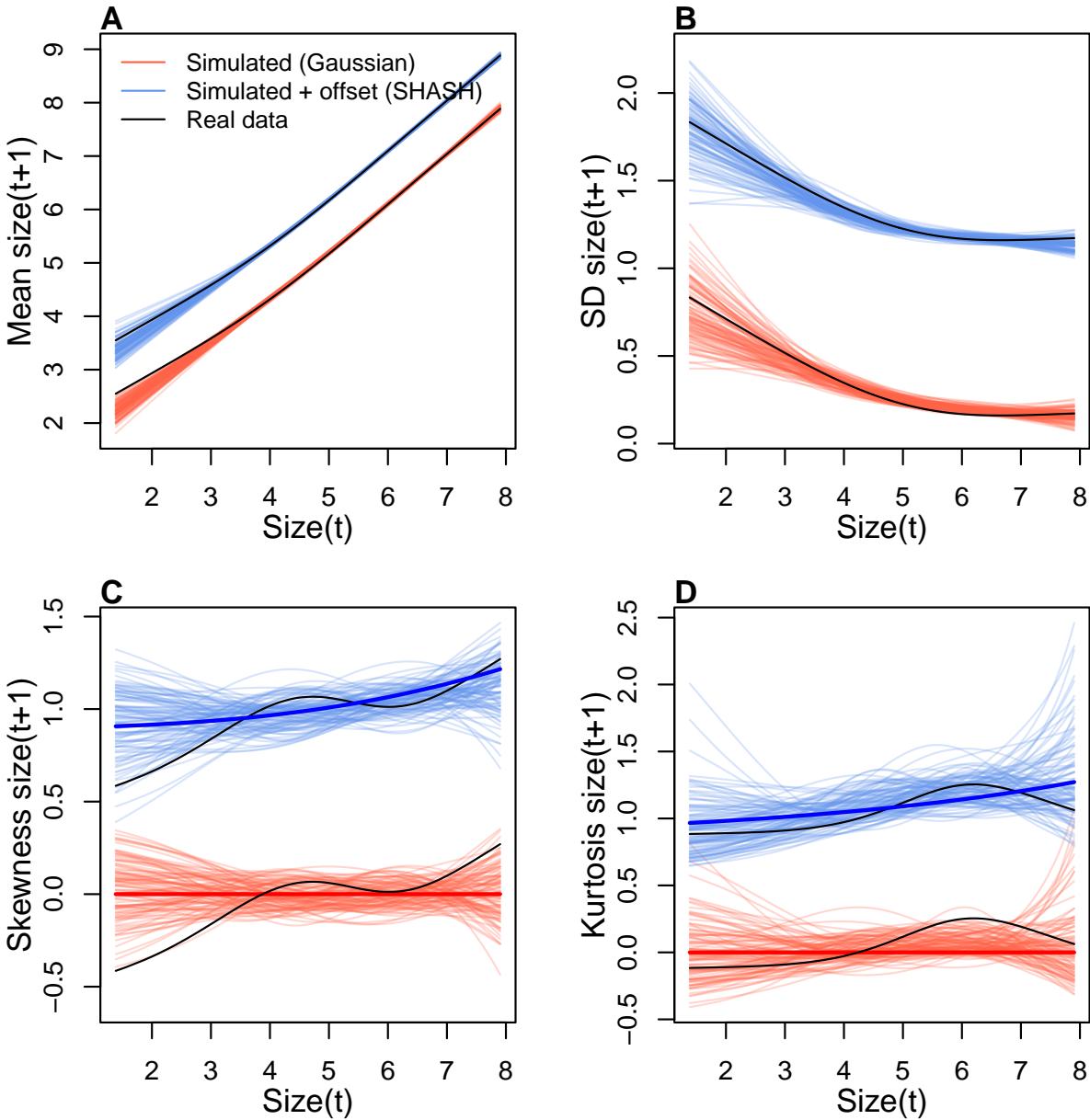


Figure S-3: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Note that plotted values for the SHASH are offset by one unit to allow comparisons. In the skewness and kurtosis panels, the darker solid curves show the values for the fitted growth models. Figure made by script `AkumalCorals_qgam.R`.

```

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

```

216 The fitted model's mean and variance are nearly identical to the Gaussian (Fig. S-3AB), and the fitted trends
 217 in skewness and kurtosis are much less “wiggly” than the estimate from the data (Fig. S-3CD). Nonetheless,

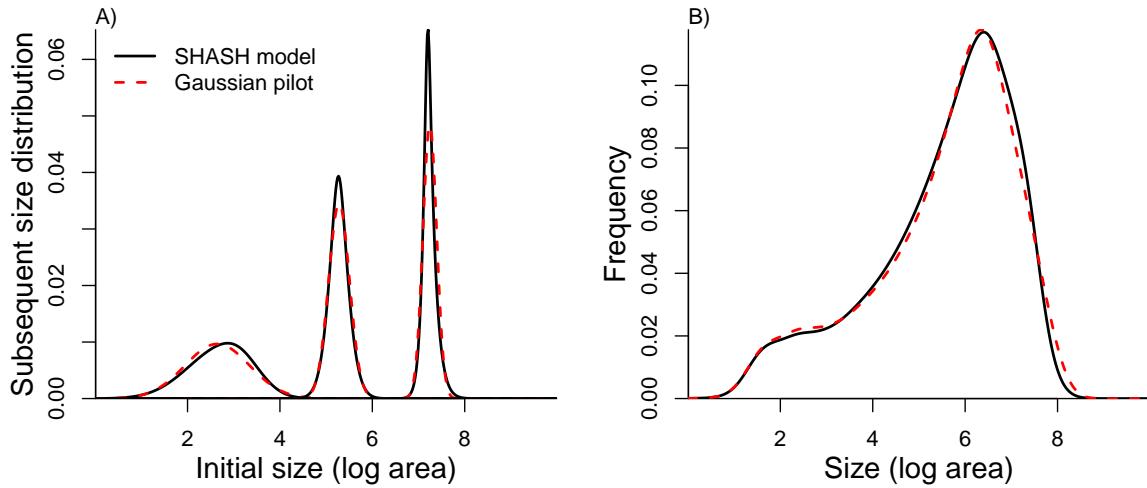


Figure S-4: Comparisons between the fitted SHASH growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year $t + 1$ for three different values of size in year t . The leftmost pair of curves are 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`.

218 data simulated from the SHASH model are more consistent with the real data, with more SHASH data
 219 sets matching or exceeding the largest skewness and kurtosis values observed (Fig. S-3CD). If one cares
 220 to quantify the difference between models, the SHASH model is clearly favored by AIC ($\Delta AIC = 5.45$)
 221 despite having twice as many parameters to fit.

222 What, then, have we gained by fitting a better growth model? Fig. S-4A compares the predicted
 223 distributions of subsequent size in the fitted model and Gaussian pilot models, for the median size of a new
 224 recruit (leftmost pair of curves), the median initial size (central curves), and the 95th percentile of initial size
 225 in the data (rightmost curves). The differences are small, and most pronounced for the smallest size, where
 226 recruits are predicted to grow slightly larger under the SHASH model than the Gaussian model. The direction
 227 of this difference was surprising, because the SHASH has negative skew at small sizes in the data. However,
 228 the SHASH model also gives a better prediction of mean growth at small sizes than the Gaussian model. At
 229 intermediate sizes the predictions are nearly identical; at large sizes the SHASH has slightly lower standard
 230 deviation, but fatter tails (excess kurtosis). Fig. S-4B shows the predicted steady-state size distributions
 231 resulting from a constant unit input of recruits. Again, the differences are very subtle. Finally, the Gaussian
 232 and SHASH growth models predict very similar mean life span (17.7 and 17.9 years, respectively).

233 In this case study we used `gam` to fit both the Gaussian and SHASH models because that obviated
 234 model selection on functions for mean, variance, and higher moments. However, `gam` should be used

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

241 S.1.2 Creosotebush, *Larrea tridentata*

242 Our next case study comes from our studies of the woody shrub creosotebush (*Larrea tridentata*) at the
243 Sevilleta Long-Term Ecological Research (LTER) site in central New Mexico, US. At this site as elsewhere
244 in the Southwest US, creosotebush is encroaching into desert grassland habitats. The data described here
245 were collected along transects spanning grass-shrub ecotones to understand patterns of density dependence
246 in creosotebush demography. Specifically, we asked whether fitness is maximized approaching zero
247 density at the leading edge of the expansion front (consistent with ‘pulled’ expansion), or whether there
248 is a demographic advantage for shrubs at higher density due to positive feedbacks expected for ecosystem
249 engineers (leading to ‘pushed’ expansion). Our published study (Drees *et al.*, 2023) used a spatial integral
250 projection model (SIPM) to predict the speed of shrub encroachment, assuming normally-distributed size
251 transitions with non-constant variance. Here we ask whether a non-Gaussian model would have been more
252 faithful to the data, and how such an improvement would influence predictions for the speed of encroachment.

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

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

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

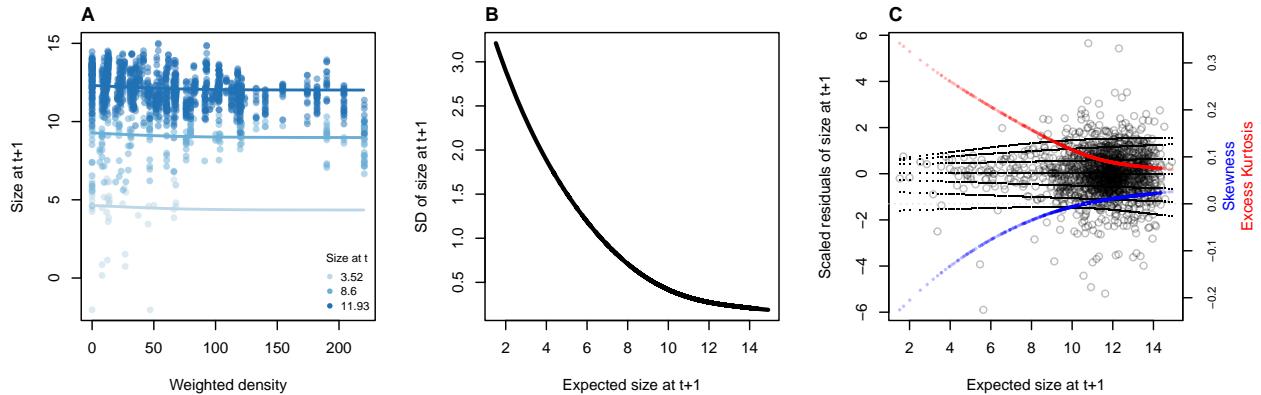


Figure S-5: **A**, Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume (cm^3). **B**, Standard deviation of size at time $t+1$ as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

Using the fitted values from this initial model, we updated the standard deviation function to be a smooth function of fitted values, and iterated the fitting until the weights stopped changing, following the same steps as in the orchid case study. As with tree cholla cactus, the standard deviation function required $k=6$ basis functions to pass our graphical diagnostic (Fig. S-1C,D). The remaining small, nearly linear trend in the scale of standardized residuals (Fig. S-1D) is not improved by using $k=8$ basis functions, and appears to be driven by high leverage points in a region of relatively sparse data, so we did not attempt to further improve the pilot model.

The resulting Gaussian growth model predicts strong initial size-dependence and weak and slightly nonlinear (but monotonic) negative density dependence (Fig. S-5A). The model indicates non-constant variance, with greater dispersion at smaller sizes (Fig. S-5B).

Quantiles of the standardized residuals indicate that skew and excess kurtosis are both greater at smaller sizes (Fig. S-5C). Skewness is close to zero for larger plants (the best-sampled size range) but excess kurtosis remains positive for large plants (ca. 10% heavier tails than Gaussian). As a candidate for improvement, we turned to the Johnson's S_U (JSU) distribution, a four-parameter, leptokurtic distribution capable of skew in either direction.

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

```

288 JSULogLik=function(pars){
289   dJSU(LATR_grow$log_volume_t1,
290   mu=LATR_grow$GAU_mean,
291   sigma=LATR_grow$GAU_sd,
292   nu = pars[1]+pars[2]*LATR_grow$GAU_mean,
293   tau = exp(pars[3]+pars[4]*LATR_grow$GAU_mean), log=TRUE)
294 }

```

295 The mean and standard deviation of the JSU are set to those of the best Gaussian model and parameters
296 controlling skewness and kurtosis were fit independently, following our approach to the orchid data. The
297 hybrid JSU model performed well, generating simulated data that aligned with the real data better than the best
298 Gaussian model, particularly in the standard deviation and kurtosis (Fig. S-6). The JSU model has exactly
299 the same mean and standard deviation of future size as the Gaussian, but Fig. S-5 uses the quantile-based
300 nonparametric mean and standard deviation. The results show that even though the JSU was not fitted to
301 match those, it comes closer than the Gaussian model as a result of accounting for the skew and kurtosis.

302 The improvement of the JSU over the Gaussian growth model, while visually satisfying, had only weak
303 influence on SIPM results. The Gaussian model slightly over-estimated the low-density growth rate, but
304 models using either Gaussian or JSU growth kernels had very similar monotonic decreases in λ with increasing
305 local density, and nearly identical wave velocities (Fig. S-7). This species has very low mortality risk once
306 established (mean remaining life expectancy of a median-sized shrub is 24,408 years) and its population growth
307 and wave expansion are limited by very low seedling recruitment ((Drees *et al.*, 2023)). Weak size-dependence
308 in survival likely explains why the improvement in growth modeling had little influence on SIPM predictions.

309 S.1.3 Case study: pike, *Esox lucius*

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

318 With no additional covariates or random effects, this is a simple growth model of final size conditional
319 on initial size. We use the natural log of length. Our first step was a Gaussian model of $\log(\text{length})$ where

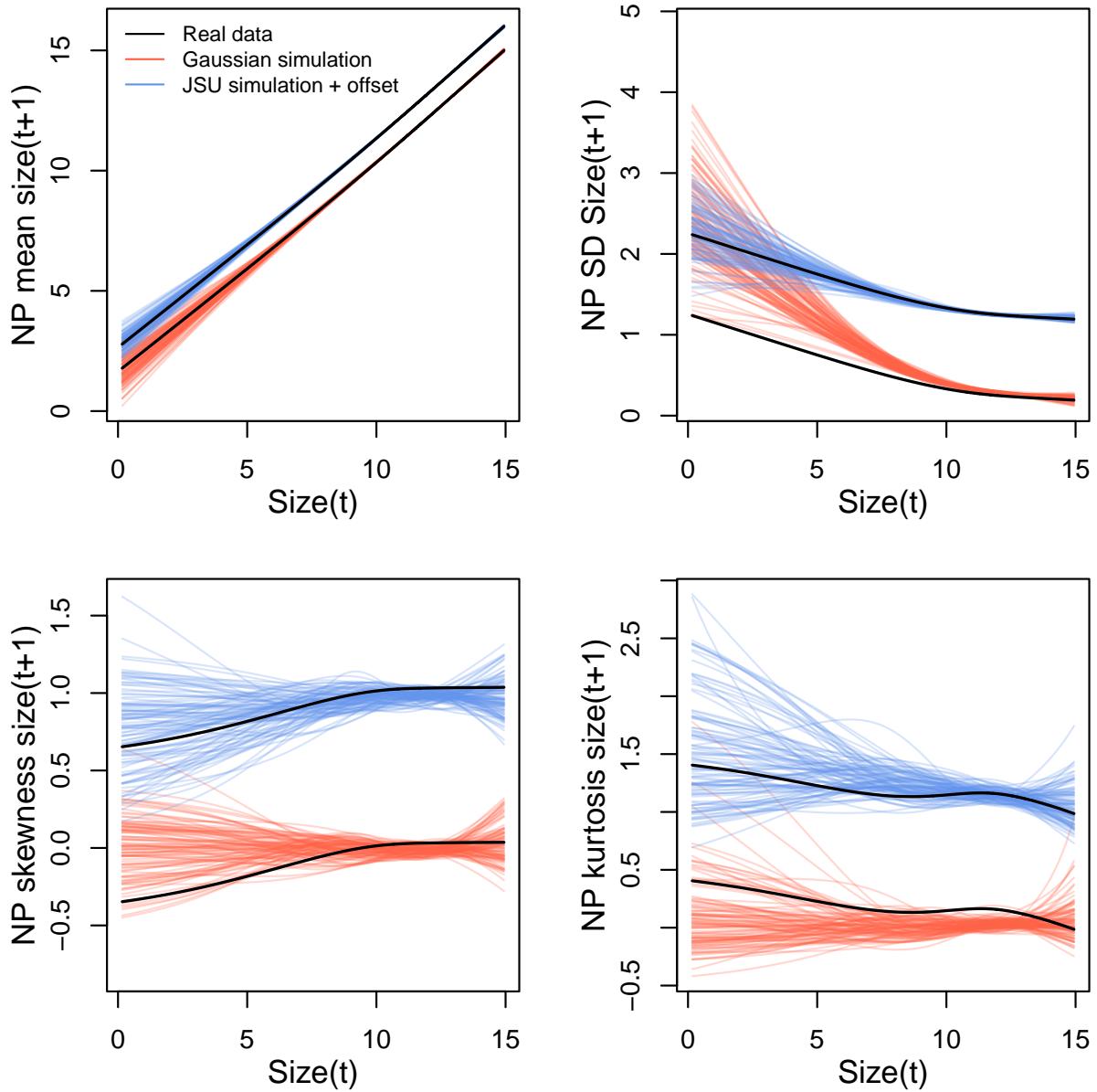


Figure S-6: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect 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`.

320 the mean and standard deviation are smooth functions of initial size fit using the `gaulss()` family in
 321 `mgcv`. We then derive the scaled residuals from the fitted mean and standard deviation:

```
322 # pike is the data frame
323 #t1 and t0 are final and initial log(length), respectively
```

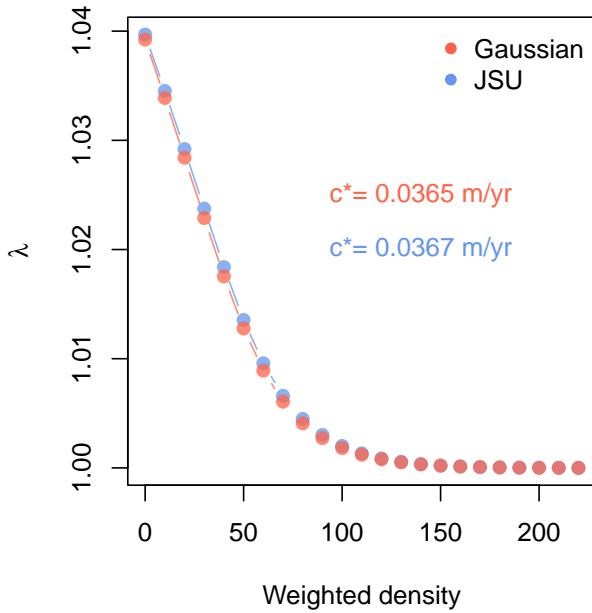


Figure S-7: Density dependence in fitness (λ) and asymptotic velocity of the creosote encroachment wave (c^*) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ($\log(cm^3)$) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

```

324 pike_gau<-gam(list(t1 ~ s(t0,k=5), ~ s(t0,k=5)), data=pike, family=gaulss())
325 pike_gau_pred<-predict(pike_gau,type="response")
326 pike$fitted_mean<-pike_gau_pred,1
327 pike$fitted_sd<-1/pike_gau_pred[,2]
328 pike$scaledResids=residuals(pike_gau,type="response")/fitted_sd

329 Based on preliminary fits we found that a basis function number of  $k = 5$  was necessary to minimize
330 variance trends in the standardized residuals. Even so, because of the very large sample size, our graphical
331 diagnostics for the pilot mean and standard deviation functions (Fig. S-1E,F) detected small-scale deviations
332 from constant mean and variance. Note that individual sizes in this study were recorded somewhat coarsely
333 (nearest 1cm). That accounts for the striking visual patterns in the scaled residuals, and probably also
334 accounts for the small-scale patterns in the diagnostic regression curves (Fig. S-1E,F). In order to remove
335 the small-scale wiggles in the diagnostic splines, we would have to introduce small-scale wiggles in the
336 mean and variance functions, which are unlikely to be real features of pike growth trajectories. So while
337 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-8B). Size transitions were fat-tailed at small initial sizes but were consistent with Gaussian tails at large initial sizes.

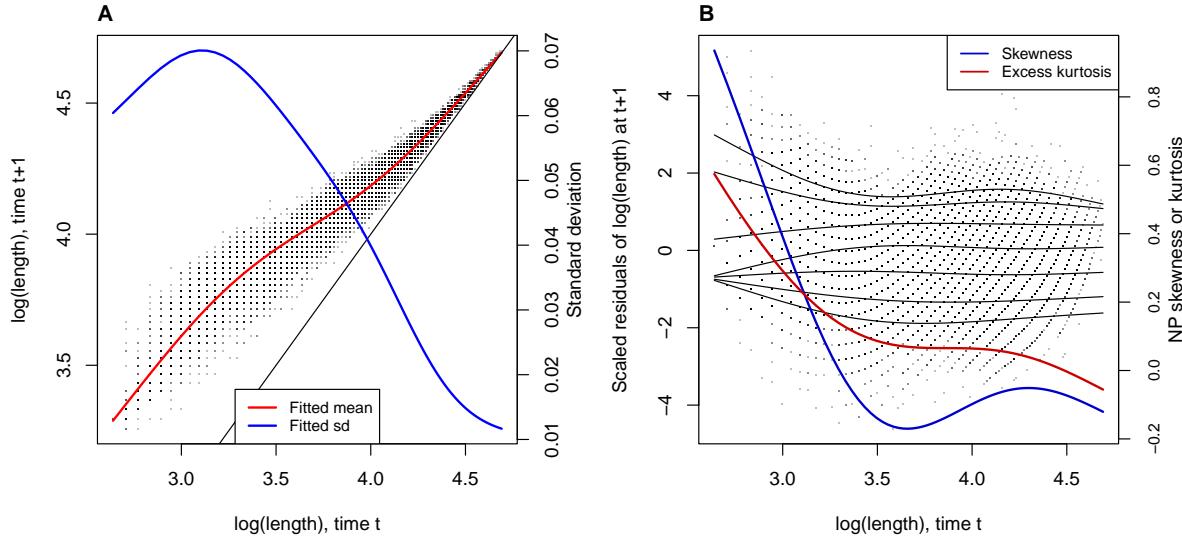


Figure S-8: **A**, Size transition data for pike, *Esox lucius*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (blue) derived from them. Black lines in **B** show the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles.

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

```
343 pike_gam_shash <- gam(list(t1 ~ s(t0,k=5), # <- model for location
344 ~ s(t0,k=5), # <- model for log-scale
345 ~ s(t0,k=5), # <- model for skewness
346 ~ s(t0,k=5)), # <- model for log-kurtosis
347 data = pike, family = shash, optimizer = "efs")
```

348 We also tried gamma regression on the change in size, to ensure strictly increasing size transitions, but found
 349 that this was not actually necessary to prevent shrinkage and did not provide as good a fit as the SHASH. Data
 350 simulated from the SHASH and Gaussian models are shown in Fig. S-9. The SHASH is an improvement
 351 over the Gaussian for most initial sizes. It fails to capture kurtosis of the largest fish, but that will have little
 352 effect because the fitted mean and standard deviation imply, correctly, that those fish will have very small and
 353 nearly deterministic growth increments until they reach the size at which growth ceases (Figs. S-8A, S-9A).

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

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

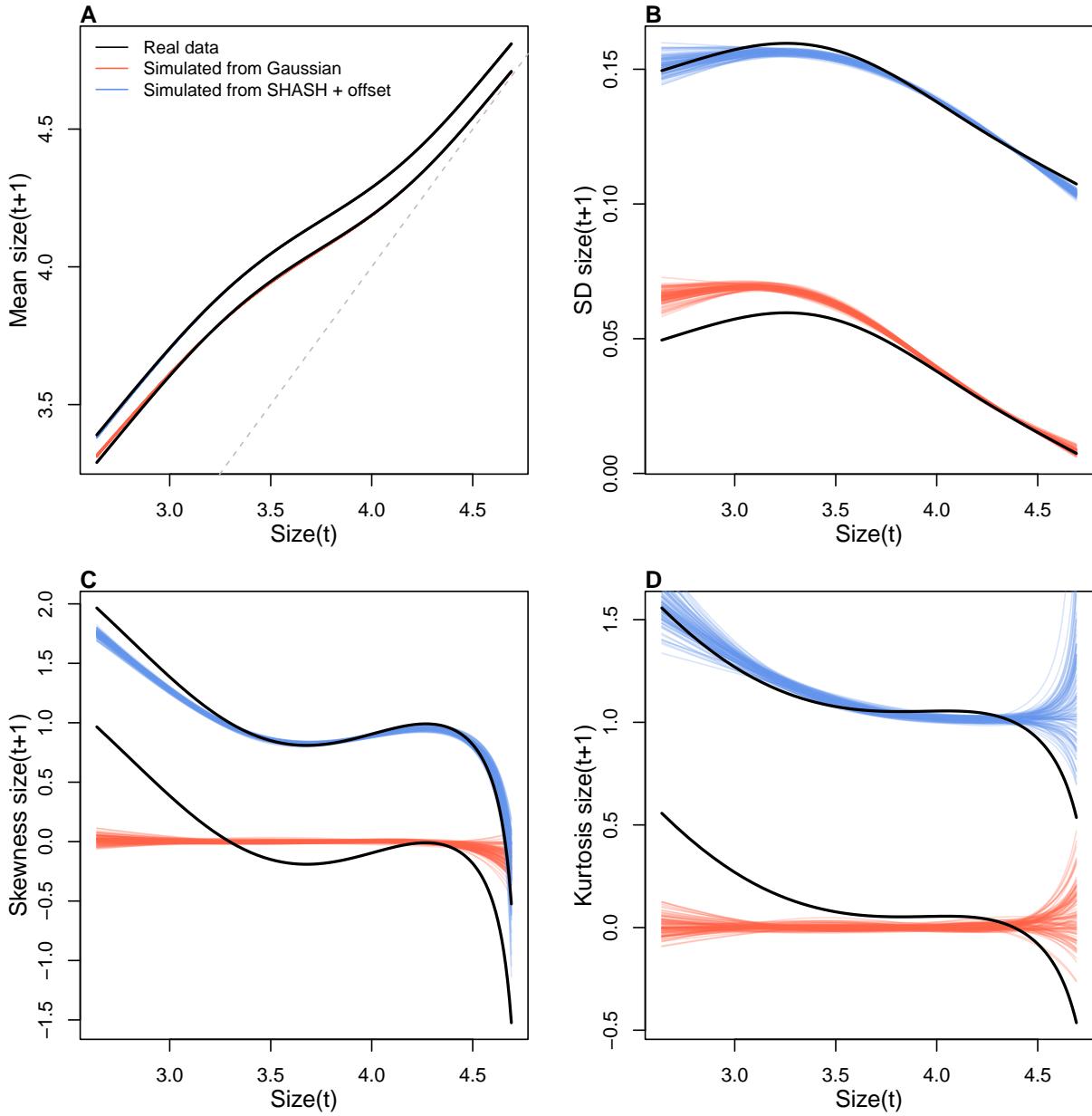


Figure S-9: Comparisons between real pike data and data simulated from Gaussian and SHASH growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect to initial size. The dashed line in the top-left panel is the 1:1 line. Figure made by script `PikeGrowthModeling_qgam.R`.

Appendix S4

358 S.1 Additional Figures

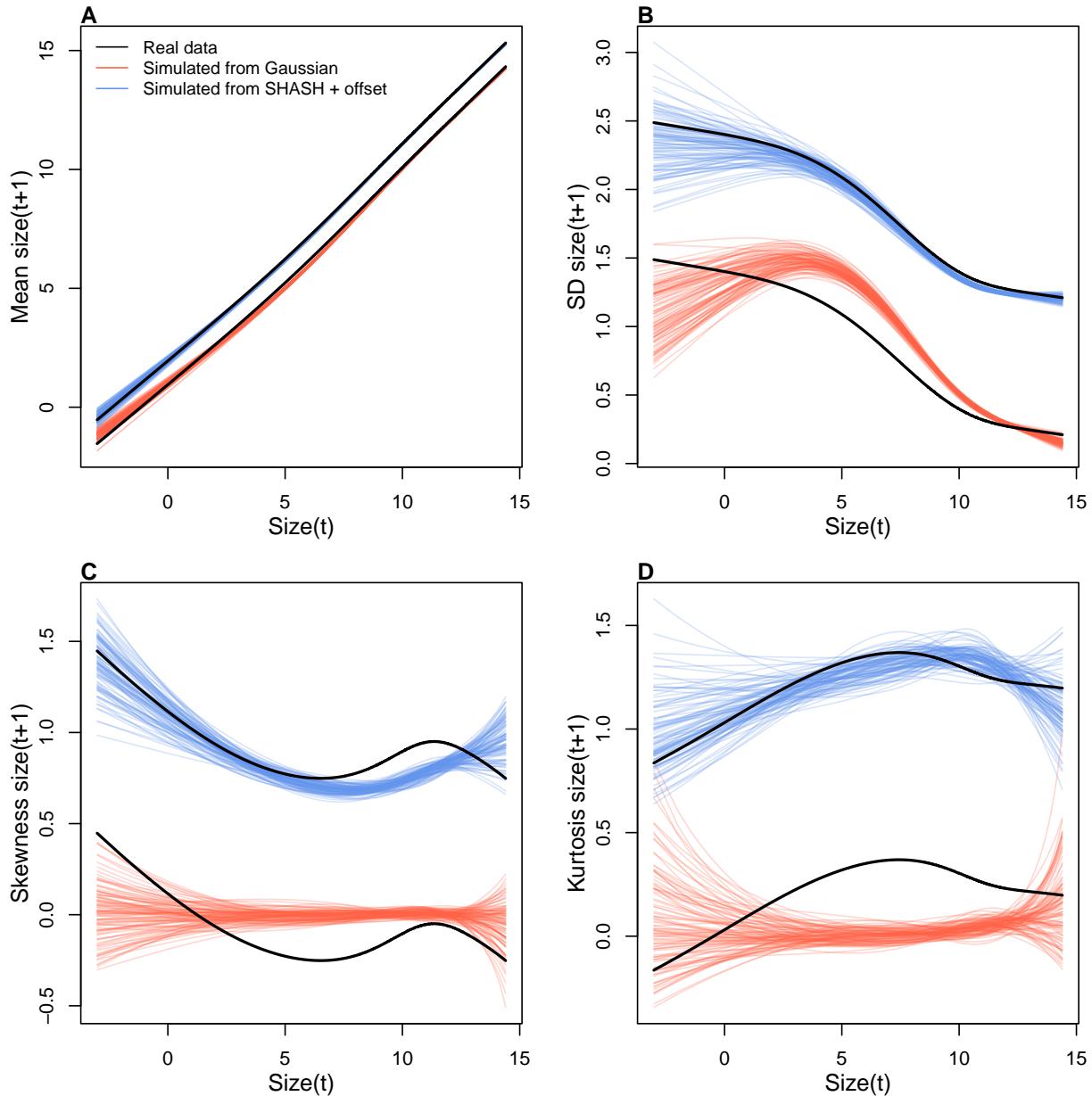


Figure S-1: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

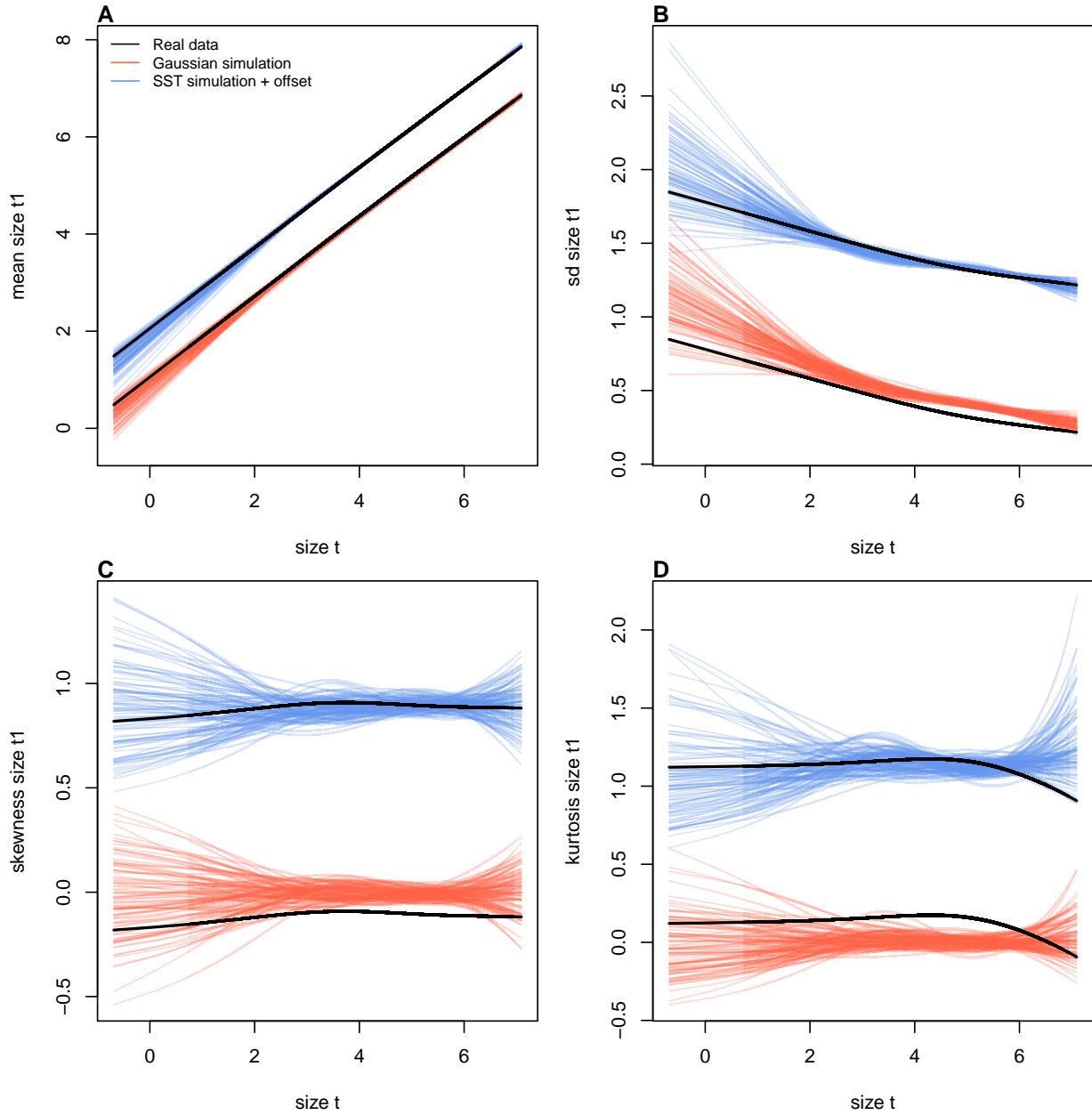


Figure S-2: Comparisons between real orchid data and data simulated from Gaussian and skewed t growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (**A-D**) shows plants that were vegetative at the start of the transition year and 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`.

359 **Literature Cited**

- 360 Abramowitz, M. & Stegun, I.A. (1970) Handbook of Mathematical Functions. 9th printing. Dover
361 Publications, Inc., New York.
- 362 Adler, P.B., Kleinhesselink, A., Hooker, G., Taylor, J.B., Teller, B. & Ellner, S.P. (2019) Data from: Weak
363 interspecific interactions in a sagebrush steppe? Conflicting evidence from observations and experiments.
364 Dryad Data set. <https://doi.org/10.5061/dryad.96dn293>.
- 365 Bruno, J.F., Ellner, S.P., Vu, I., Kim, K. & Harvell, C.D. (2011) Impacts of aspergillosis on sea fan coral
366 demography: modeling a moving target. *Ecological Monographs* **81**, 123–139.
- 367 Cooch, E.G. & White, G.C. (2020, accessed 5/17/2020) *Program MARK - a 'gentle introduction'*. phidot.org.
- 368 Drees, T., Ochocki, B.M., Collins, S.L. & Miller, T.E. (2023) Demography and dispersal at a grass-shrub eco-
369 tone: a spatial integral projection model for woody plant encroachment. *Ecological Monographs* p. e1574.
- 370 Gould, W.R. & Nichols, J.D. (1998) Estimation of temporal variability of survival in animal populations.
371 *Ecology* **79**, 2531 – 2538.
- 372 Gu, C. (2013) *Smoothing Spline ANOVA Models*. Springer Science+Business Media, New York, 2 edn.
- 373 Jones, M. & Pewsey, A. (2009) Sinh-arcsinh distributions. *Biometrika* **96**, 761 – 780.
- 374 Jones, M.C., Rosco, J.F. & Pewsey, A. (2011) Skewness-invariant measures of kurtosis. *The American
375 Statistician* **65**, 89 – 95.
- 376 Link, W.A. & Nichols, J.D. (1994) On the importance of sampling variance to investigations of temporal
377 variation in animal population size. *Oikos* **69**, 539 – 544.
- 378 Metcalf, C.J.E., Ellner, S.P., Childs, D.Z., Salguero-Gómez, R., Merow, C., McMahon, S.M., Jongejans,
379 E. & Rees, M. (2015) Statistical modelling of annual variation for inference on stochastic population
380 dynamics using Integral Projection Models. *Methods in Ecology and Evolution* **6**, 1007–1017.
- 381 Ochocki, B.M., Drees, T. & Miller, T.E. (2023) Density-dependent demography of creosote
382 bush (*larrea tridentata*) along grass-shrub ecotones. [https://doi.org/10.6073/pasta/
383 ca53c16f16dcf9fb11f3ee99ea5445ac](https://doi.org/10.6073/pasta/ca53c16f16dcf9fb11f3ee99ea5445ac).
- 384 Rigby, R.A., Stasinopoulos, M.D., Heller, G.Z. & De Bastiani, F. (2019) *Distributions for modeling location,
385 scale, and shape: Using GAMMSS in R*. CRC press.

- 386 Stubberud, M.W., Vindenes, Y., Vøllestad, L.A., Winfield, I.J., Stenseth, N.C. & Langangen, Ø. (2019)
387 Effects of size-and sex-selective harvesting: An integral projection model approach. *Ecology and*
388 *Evolution* **9**, 12556–12570.
- 389 Vindenes, Y., Edeline, E., Ohlberger, J., Langangen, Ø., Winfield, I.J., Stenseth, N.C. & Vøllestad, L.A.
390 (2014) Effects of climate change on trait-based dynamics of a top predator in freshwater ecosystems.
391 *The American Naturalist* **183**, 243–256.
- 392 Wood, S. (2017) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC, 2 edn.
- 393 Winfield, I., Fletcher, J. & James, J. (2013a) Pike fecundity data 1963-2002. NERC Environmental Information
394 Data Centre, <https://doi.org/10.5285/b8886915-14cb-44df-86fa-7ab718acf49a>.
- 395 Winfield, I., Fletcher, J. & James, J. (2013b) Pike growth data 1944-1995. NERC Environmental Information
396 Data Centre, <https://doi.org/10.5285/637d60d6-1571-49af-93f7-24c1279d884d>.
- 397 Winfield, I., Fletcher, J. & James, J. (2013c) Pike survival data 1953-1990. NERC Environmental Information
398 Data Centre, <https://doi.org/10.5285/813e07dd-2135-49bc-93c6-83999e442b36>.