

Supporting Information

616 1 The Jones-Pewsey SHASH distribution family

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

$$622 Z = \sinh(\delta \text{asinh}(X_{\varepsilon,\delta}) - \varepsilon) \quad (1)$$

623 and Z has a $\text{Normal}(0,1)$ distribution. Equivalently,

$$624 X_{\varepsilon,\delta} = \sinh(\delta^{-1}(\text{asinh}(Z) + \varepsilon)), \quad Z \sim \mathcal{N}(0,1). \quad (2)$$

625 Parameters $\delta = 1, \varepsilon = 0$ give the $\text{Normal}(0,1)$ distribution. Skewness has the sign of ε , and $\delta > 0$ controls
626 tail weight, with heavier than Gaussian tails for $\delta < 1$ and lighter than Gaussian tails for $\delta > 1$. A formula
627 for the density $f_{\varepsilon,\delta}$ is given by Jones & Pewsey (2009, eqn. 2). The general four-parameter distribution with
628 location parameter μ and scale parameter σ is defined as the probability density of $\mu + \sigma X_{\varepsilon,\delta}$. This is often
629 called the SHASH distribution.

630 As is unfortunately the case for most four-parameter distributions, μ is not the mean, σ is not the
631 standard deviation, ε is not the skew and δ is not the kurtosis. All else being equal, larger μ gives a larger
632 mean, larger σ gives a higher standard deviation, higher ε gives higher asymmetry, and smaller δ gives
633 heavier tail weight. But all moments are jointly determined by all four parameters.

634 The countervailing advantage of the SHASH family is that the attainable combinations of skewness
635 and kurtosis are very broad compared to other four-parameter families, and come very close to the theoretical
636 limit of kurtosis as a function of skewness (Jones & Pewsey, 2009, Fig. 2). Additionally, eqn. (2) makes it
637 straightforward to generate random numbers and to compute the probability density, cumulative distribution,
638 and quantile functions. There are also analytic formulas for the first four moments (Jones & Pewsey, 2009,
639 p. 764). These make it possible to define a reparameterized form of the distribution in which the location
640 parameter is the mean, and the scale parameter is the standard deviation. That form can then be used in
641 custom likelihood functions that “import” the fitted mean and standard deviation from the Gaussian pilot
642 model, in the same way that the skewed t distribution was used in our lady orchid case study. Code for
643 this reparameterized form of the distribution is provided in our code archive script `JPfuns.R`.

644 It eventually transpired that none of our case studies involved fitting a SHASH distribution through
645 a “custom” likelihood function (though this was transiently the case). But for the sake of possible future
646 applications, we record here some advice for doing so based on our experience.

647 Eqn. (2) shows that the distribution depends on ε only through the ratio ε/δ . We have found that
648 this property can be problematic for parameter estimation, because of a ridge in the likelihood surface
649 with constant ε/δ . Another problem is that when δ is large, changes in ε have little effect.

650 To avoid those problems, we recommend writing a custom likelihood function in terms of skewness
651 and kurtosis parameters λ and τ , setting $\delta = e^\tau$, $\varepsilon = e^\tau \lambda$ in the reparameterized SHASH distribution. λ
652 can take any real value, and the distribution’s skew has the same sign as λ . τ also can take any real value,
653 with negative values giving thinner than Gaussian tails and positive values giving fatter than Gaussian tails.

654 It is still the case that skewness and kurtosis jointly depend on τ and λ , but the “crosstalk” is weaker
655 than that between δ and ε . As a result, we found that likelihood optimization is numerically more stable
656 when the likelihood function is written as a function of τ and λ rather than δ and ε , in addition to μ and σ .

657 2 Estimating random effects in non-Gaussian models using shrinkage

658 Specialized software for fitting mixed effects models only allow a subset, usually a small subset, of the
659 distributions that are useful for modeling growth.⁴ One way past this limitation is Bayesian estimation.
660 Here we describe another option, introduced by Link & Nichols (1994) and Gould & Nichols (1998): fitting
661 the model in a fixed effects framework by Maximum Likelihood, followed by shrinkage of coefficient
662 estimates. None of the ideas here are original. This section overlaps Appendix S1 of Metcalf *et al.* (2015),
663 the only new wrinkle being the application to non-Gaussian models.

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

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

⁴The `gamlss` package includes many distributions, but in our experience even with simple random effects structure the fitting algorithms often fail to converge reliably.

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

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

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

```
678 LogLik=function(pars,new.size,U){  
679     pars1 = pars[1:ncol(U)]; pars2=pars[-(1:ncol(U))];  
680     mu = U%*%pars1;  
681     sigma = exp(pars2[1]+pars2[2]*mu);  
682     dSN1(new.size, mu=mu, sigma=sigma, nu=pars2[3], log=TRUE)  
683 }
```

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

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

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

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

$$\hat{a}_t = a_t + \varepsilon_t \quad (3)$$

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

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

$$\mathbb{E}(\varepsilon_t | a_t) = 0. \quad (4)$$

702 We also adopt the standard mixed-model assumption that the a_t are drawn independently from some
703 fixed distribution with unknown variance σ^2 . These are optimistic assumptions, but not excessively so.

704 If the assumptions of maximum likelihood are satisfied, the bias in parameter estimates is asymptotically
 705 negligible compared to the standard error. The terms resulting from non-independence can only be reliably
 706 estimated if the autocorrelations fall to nearly zero within lag $m \ll T$, and in that case the autocorrelation
 707 correction term is small (see eqn. (1) in Gould & Nichols (1998)). We therefore recommend proceeding
 708 on the assumption that the \hat{a}_t are independent.

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

$$710 \quad \mathbb{E}(S^2) = \sigma^2 + \frac{1}{T} \sum_{t=1}^T \mathbb{E}Var(\varepsilon_t) - \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \mathbb{E}Cov(\varepsilon_i, \varepsilon_j). \quad (5)$$

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

713 The terms besides σ^2 on the right-hand of (5) makes S^2 a biased estimate of σ^2 . However, those
 714 terms correspond to entries in the variance-covariance matrix V , so we can use \hat{V} to remove the bias:

$$715 \quad \hat{\sigma}^2 = S^2 - \frac{1}{T} \sum_{t=1}^T \hat{V}_{t,t} + \frac{1}{T(T-1)} \sum_{i=1}^{j-1} \sum_{j=1}^T \hat{V}_{i,j}. \quad (6)$$

716 $\hat{\sigma}^2$ is the estimated variance of the distribution from which the a_t are assumed to be drawn.

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

$$720 \quad \tilde{a}_t = \bar{\hat{a}}_t + \sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{V}_{t,t}}} (\hat{a}_t - \bar{\hat{a}}_t). \quad (7)$$

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

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

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

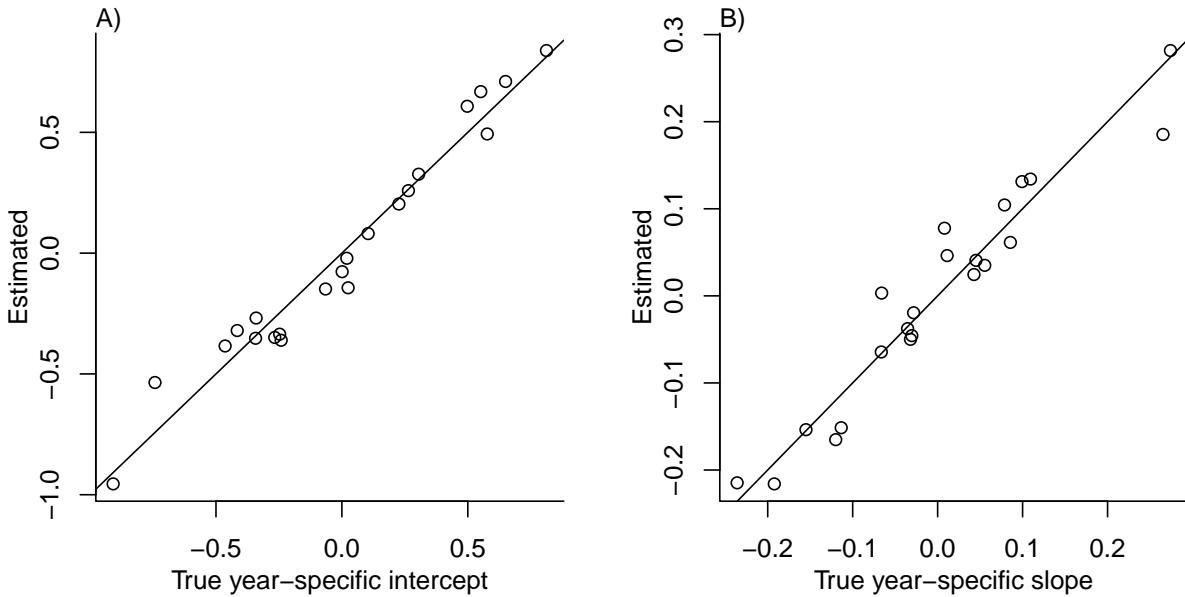


Figure 1: Comparison of the true random year effects with the shrinkage estimates, for one artificial data set generated from the fitted growth model for *Pseudoroegneria spicata*. Figure made by R script `SimpleShrinkageExample.R` in our code archive.

```

731
732 # Estimate sigma^2
733 var.hat = mean(fixed.fx^2) - mean(diag(V1)) +
734   (sum(V1)-sum(diag(V1)))/(2*T*(T-1));
735
736 # Shrink deviations from the mean
737 shrinkRanIntercept = fixed.fx*sqrt(var.hat/(var.hat + diag(V1)));
738
739 # Do it all again for the slopes
740 fixed.fx2 = coefs[(T+1):(2*T)]; fixed.fx2 = fixed.fx2-mean(fixed.fx2);
741 var2.hat = mean(fixed.fx2^2) - mean(diag(V2)) +
742   (sum(V2)-sum(diag(V2)))/(2*T*(T-1));
743 shrinkRanSlope = fixed.fx2*sqrt(var2.hat/(var2.hat + diag(V2)));

```

Figure 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.

748 **3 Additional case studies**

749 **3.1 Sea fan corals, *Gorgonia ventalina***

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

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

```
765 # XH is a data frame holding the data  
766 # logarea.t0, .t1 denote initial and final values of log-transformed area  
767 fitGAU <- gam(list(logarea.t1~s(logarea.t0), ~s(logarea.t0)),  
768 data=XH, gamma=1.4, family=gaulss())  
769 fitted_all = predict(fitGAU,type="response");  
770 fitted_sd = 1/fitted_all[,2];  
771 scaledResids = residuals(fitGAU,type='response')/fitted_sd;
```

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

777 While there are no blatant signs of trouble in the pilot Gaussian model, quantile regressions on the scaled
778 residuals, and the NP Skewness and Kurtosis metrics derived from them (Eq. 2 and 3), suggest deviations
779 from normality (Fig. 3B). Specifically, skewness switches from negative to positive across the size range, with

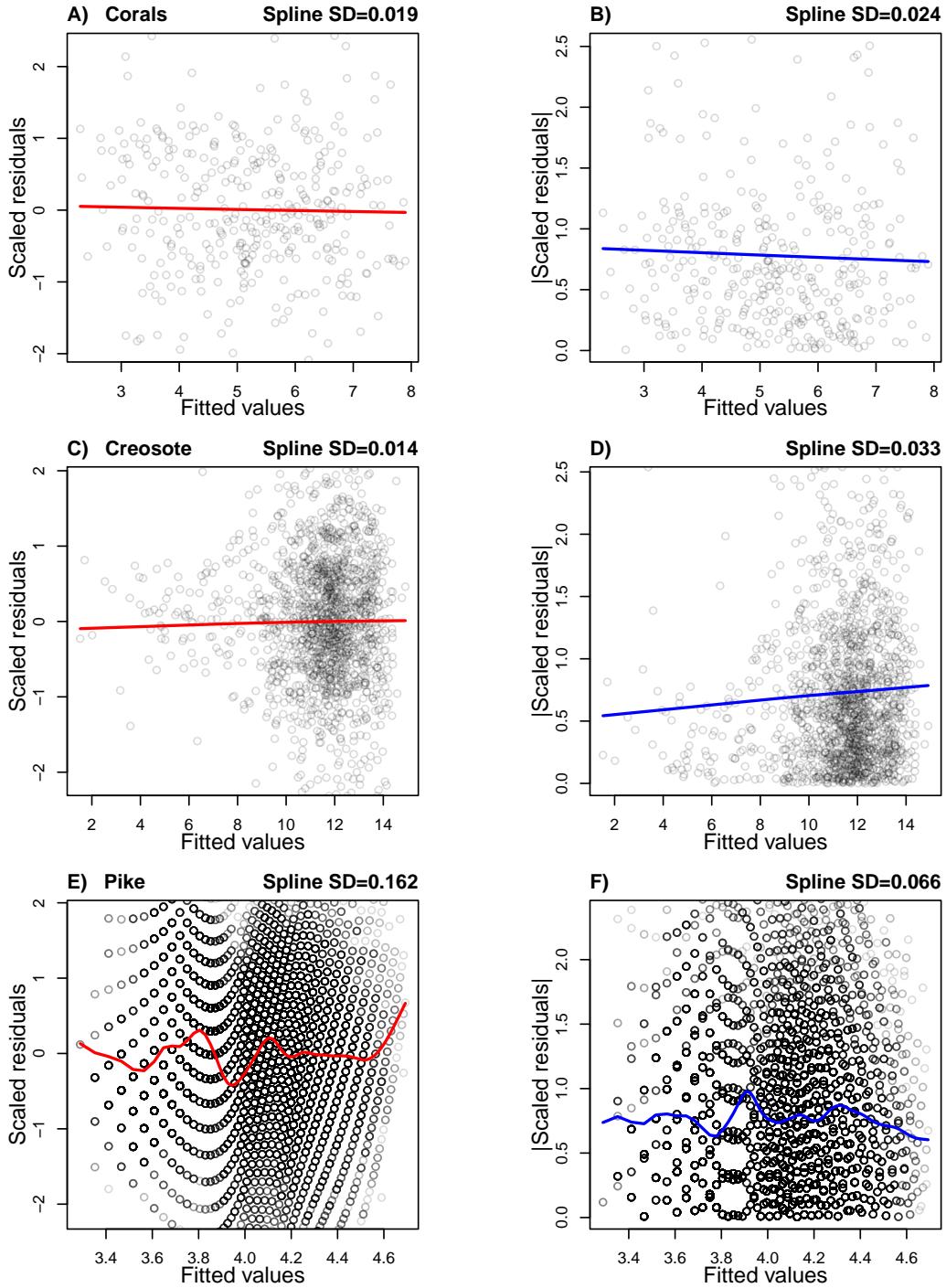


Figure 2: Diagnostic plot for trends in the mean (left column) or variance (right column) of scaled residuals from a pilot Gaussian model, for the sea fan corals A,B, creosote bush C,D, and pike E,F. In A,C,E the standardized residuals are plotted, and in B,D,F the absolute values of standardized residuals, 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`.

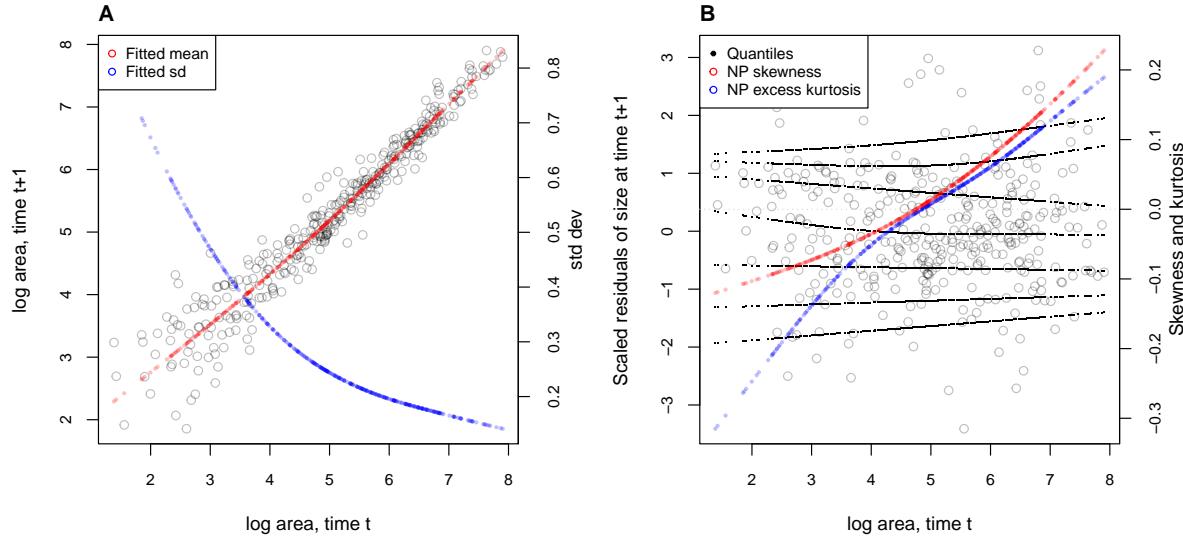


Figure 3: **A**, Size transition data for sea fan corals, *Gorgonia ventalina*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (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`.

780 smaller corals more prone to extreme shrinkage and larger corals more prone to extreme growth. Kurtosis also
 781 changes direction over the size distribution, with thinner tails than Gaussian at small sizes and fatter tails at
 782 large sizes. The fitted nonparametric moments suggest that the upper and lower tails of size transition probabili-
 783 ties may differ by up to 20%, and the weight of the tails may be 20% greater or less than Gaussian, depending
 784 on initial size – not overwhelming deficiencies, but not trivial either. Are these deviations from normality
 785 severe enough to warrant a second, non-Gaussian iteration of growth modeling? To answer that question, we
 786 simulated data from the fitted Gaussian model and examined whether key properties of the simulated data
 787 are consistent with those of the real data. If the simulated data are not consistent with the real data, it is time
 788 to choose a better distribution (Fig. 1). In this case, most of 100 Gaussian model simulations are out of line
 789 with the skew at smallest and largest sizes, and excess kurtosis observed at moderately large sizes (Fig. 4 CD).
 790 For at least some parts of the size distribution, a non-Gaussian model would better capture size transitions.

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

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

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

797 ~ s(logarea.t0,k=4), # <- skewness
```

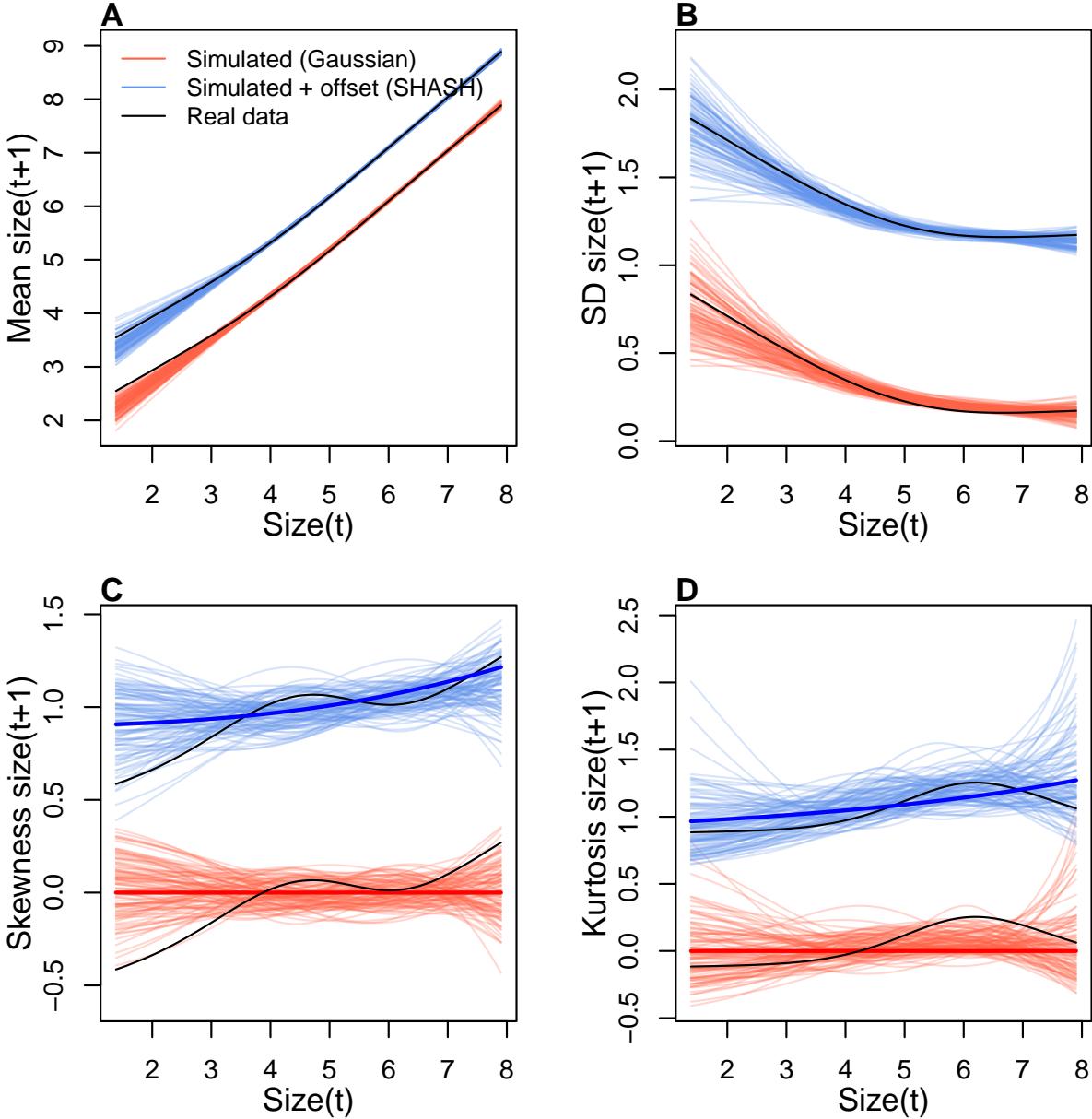


Figure 4: Comparisons among real coral data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Note that plotted values for the SHASH are offset by one unit to allow comparisons. In the skewness and kurtosis panels, the darker solid curves show the values for the fitted growth models. Figure made by script `AkumalCorals_qgam.R`.

```

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

```

800 The fitted model's mean and variance are nearly identical to the Gaussian (Fig. 4AB), and the fitted trends
 801 in skewness and kurtosis are much less “wiggly” than the estimate from the data (Fig. 4CD). Nonetheless,
 802 data simulated from the SHASH model are more consistent with the real data, with more SHASH data

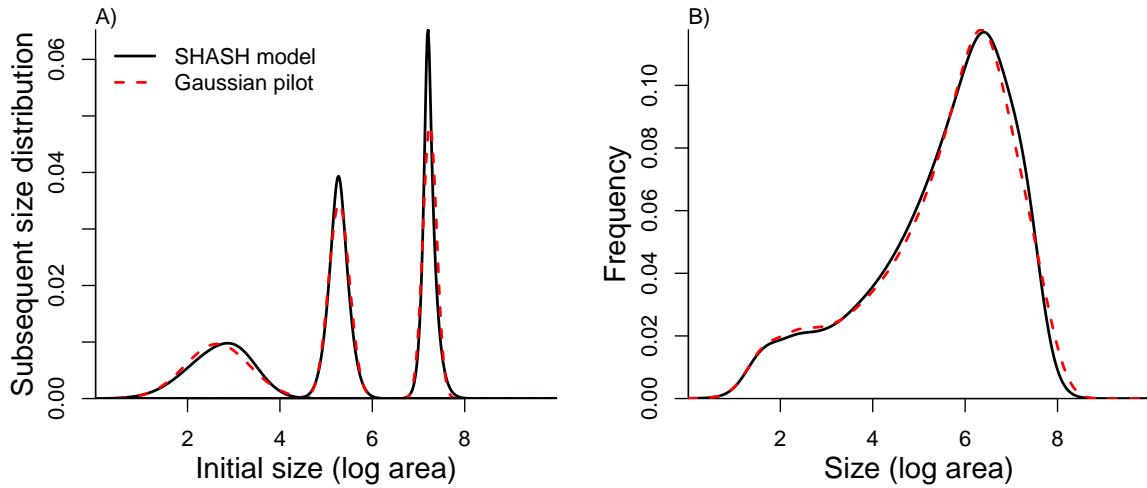


Figure 5: Comparisons between the fitted SHASH growth model (solid black curves) and the Gaussian pilot model (dashed red curves) for sea fans *G. ventalina*. A) Predicted frequency distributions of size in year $t + 1$ for three different values of size in year t . The leftmost pair of curves are for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

sets matching or exceeding the largest skewness and kurtosis values observed (Fig. 4CD). If one cares to quantify the difference between models, the SHASH model is clearly favored by AIC ($\Delta AIC = 5.45$) despite having twice as many parameters to fit.

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

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

825 **3.2 Creosotebush, *Larrea tridentata***

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

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

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

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

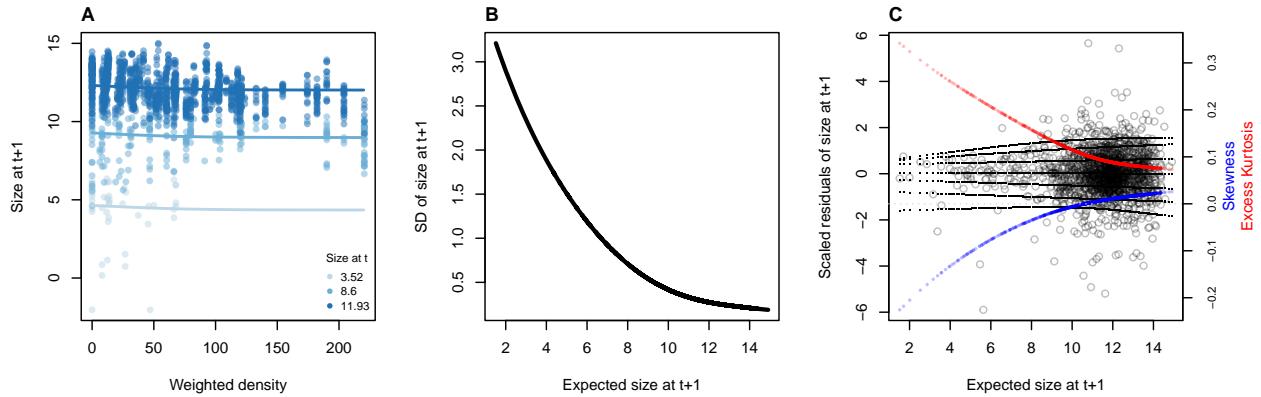


Figure 6: **A.** Creosotebush size transition data with respect to initial size (colors) and local weighted density (sum of sizes of all plants within a five-meter transect window). Size is quantified as the natural logarithm of plant volume (cm^3). **B.** Standard deviation of size at time $t + 1$ as 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. 2C,D). The remaining small, nearly linear trend in the scale of standardized residuals (Fig. 2D) 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. 6A). The model indicates non-constant variance, with greater dispersion at smaller sizes (Fig. 6B).

Quantiles of the standardized residuals indicate that skew and excess kurtosis are both greater at smaller sizes (Fig. 6C). 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

870 functions from Gaussian pilot model, and estimates the parameters that control skewness and kurtosis
871 as linear functions of predicted future size. The “hybrid” likelihood looks like this:

```
872 JSULogLik=function(pars){  
873     dJSU(LATR_grow$log_volume_t1,  
874     mu=LATR_grow$GAU_mean,  
875     sigma=LATR_grow$GAU_sd,  
876     nu = pars[1]+pars[2]*LATR_grow$GAU_mean,  
877     tau = exp(pars[3]+pars[4]*LATR_grow$GAU_mean), log=TRUE)  
878 }
```

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

886 The improvement of the JSU over the Gaussian growth model, while visually satisfying, had only weak
887 influence on SIPM results. The Gaussian model slightly over-estimated the low-density growth rate, but mod-
888 els using either Gaussian or JSU growth kernels had very similar monotonic decreases in λ with increasing
889 local density, and nearly identical wave velocities (Fig. 8). This species has very low mortality risk once estab-
890 lished (mean remaining life expectancy of a median-sized shrub is 24,408 years) and its population growth and
891 wave expansion are limited by very low seedling recruitment ((Drees *et al.*, 2023)). Weak size-dependence in
892 survival likely explains why the improvement in growth modeling had little influence on SIPM predictions.

893 3.3 Case study: pike, *Esox lucius*

894 Our final case study comes from a long-term (51 year) study of pike (*Esox lucius*) at Windemere in the
895 English Lake District, UK. Fish were gill-netted and destructively sampled to retrieve otoliths. Lengths
896 (cm) were recorded at the time of sampling and back-casted to estimate length in the preceding year. There
897 were 26501 size transitions in the data set. These data are publicly available (Winfield *et al.*, 2013b), as are
898 data on size-specific fertility and survival (Winfield *et al.*, 2013a,c), and have been analyzed in previous IPM
899 studies (Vindenes *et al.*, 2014; Stubberud *et al.*, 2019). Previous authors modeled growth using a log-normal
900 distribution to ensure that change in length was non-negative. Here, we do not attempt to reproduce the

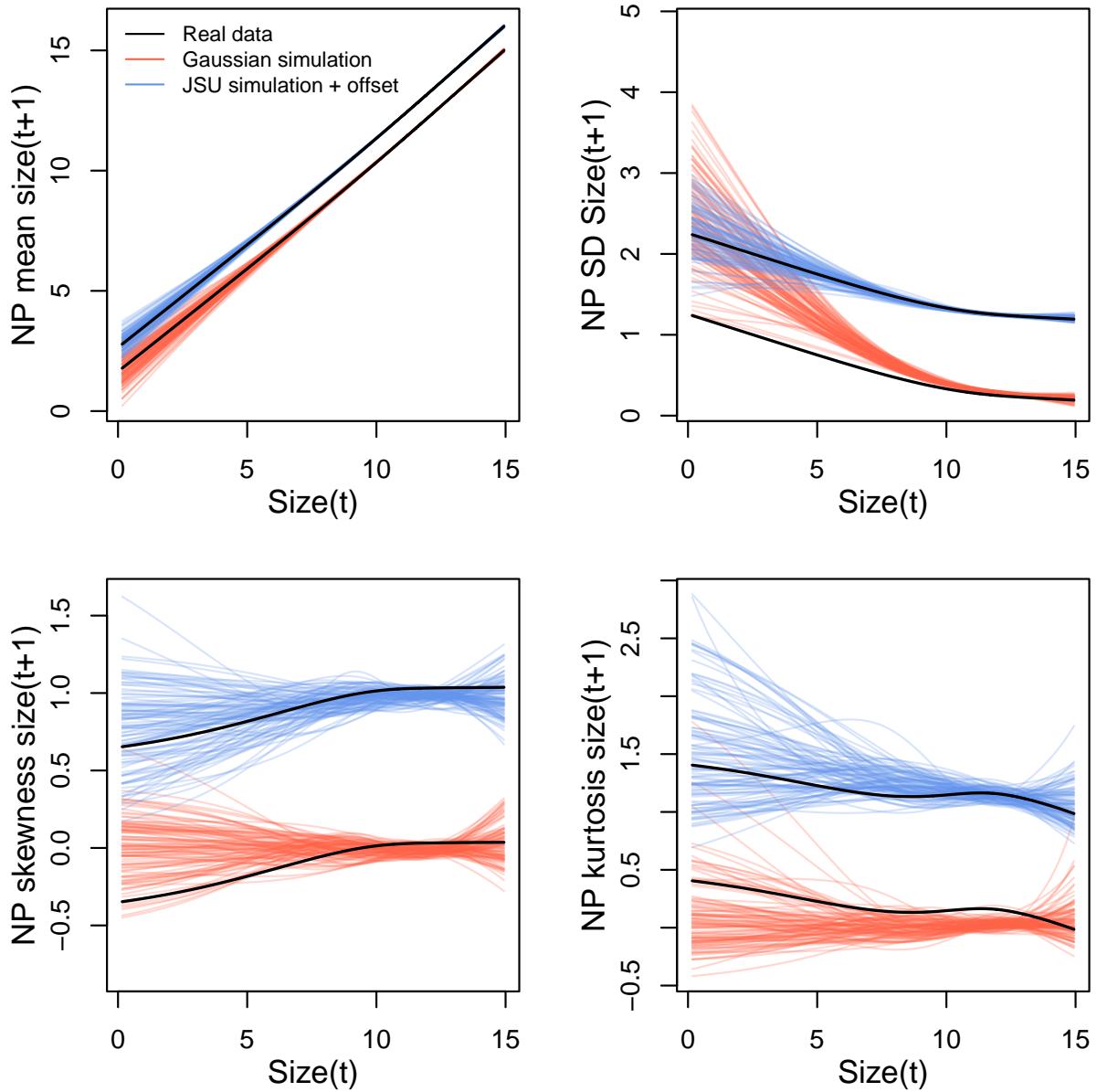


Figure 7: Comparisons between real creosotebush data and data simulated from Gaussian and JSU growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect 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`.

901 published IPMs but rather use the growth data as an additional test case of non-Gaussian growth modeling
 902 for a short-lived vertebrate.

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

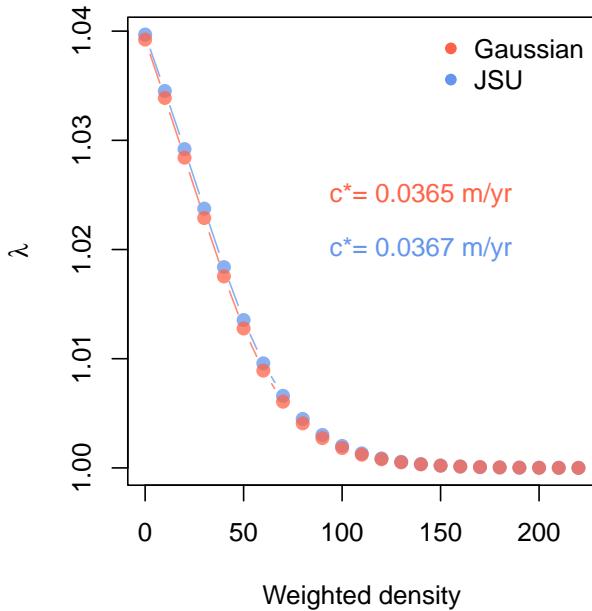


Figure 8: Density dependence in fitness (λ) and asymptotic velocity of the creosote encroachment wave (c^*) for Gaussian and JSU growth kernels. Weighted density is the sum of sizes ($\log(cm^3)$) of all conspecifics within a five-meter transect “window”. Figure made by script `creosote_growth_modeling_qgam.R`.

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

```
907 # pike is the data frame
908 #t1 and t0 are final and initial log(length), respectively
909 pike_gau<-gam(list(t1 ~ s(t0,k=5), ~ s(t0,k=5)), data=pike, family=gaulss())
910 pike_gau_pred<-predict(pike_gau,type="response")
911 pike$fitted_mean<-pike_gau_pred[,1]
912 pike$fitted_sd<-1/pike_gau_pred[,2]
913 pike$scaledResids=residuals(pike_gau,type="response")/fitted_sd
```

914 Based on preliminary fits we found that a basis function number of $k = 5$ was necessary to minimize
 915 variance trends in the standardized residuals. Even so, because of the very large sample size, our graphical
 916 diagnostics for the pilot mean and standard deviation functions (Fig. 2E,F) detected small-scale deviations
 917 from constant mean and variance. Note that individual sizes in this study were recorded somewhat coarsely
 918 (nearest 1cm). That accounts for the striking visual patterns in the scaled residuals, and probably also
 919 accounts for the small-scale patterns in the diagnostic regression curves (Fig. 2E,F). In order to remove
 920 the small-scale wiggles in the diagnostic splines, we would have to introduce small-scale wiggles in the

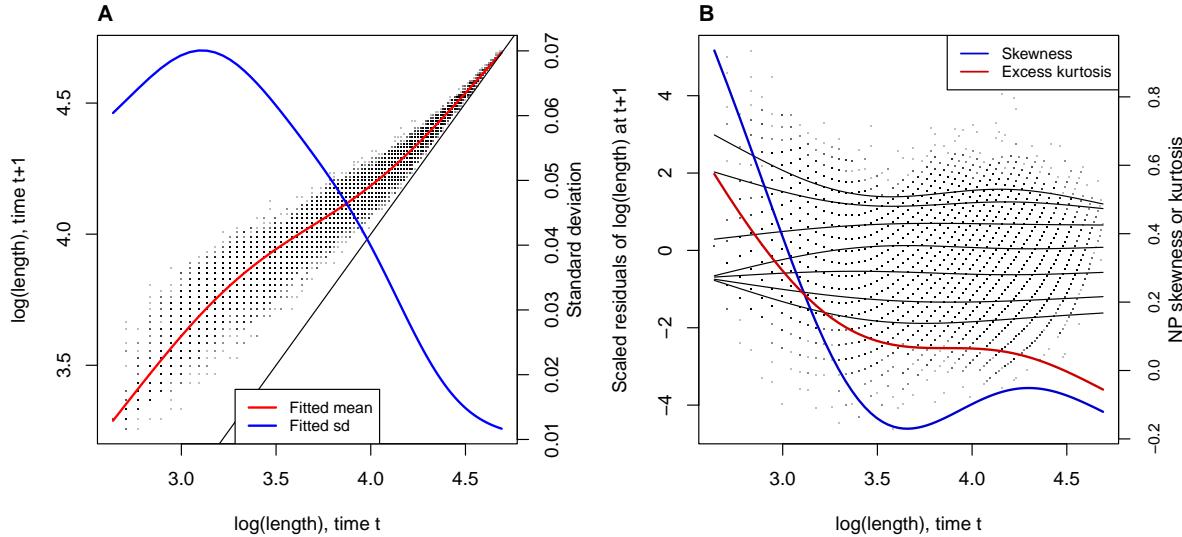


Figure 9: **A**, Size transition data for pike, *Esox lucius*, and fitted gam with mean (red) and standard deviation (blue) of future size conditional on current size. **B**, Quantile regressions of scaled residuals on size and nonparametric estimates of skewness (red) and excess kurtosis (blue) derived from them. Black lines in **B** show the 5th, 10th, 25th, 50th, 75th, 90th, and 95th quantiles.

921 mean and variance functions, which are unlikely to be real features of pike growth trajectories. So while
922 we are not entirely satisfied with our pilot model, we see no way to improve it.

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

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

```
928 pike_gam_shash <- gam(list(t1 ~ s(t0,k=5), # <- model for location
929   ~ s(t0,k=5), # <- model for log-scale
930   ~ s(t0,k=5), # <- model for skewness
931   ~ s(t0,k=5)), # <- model for log-kurtosis
932   data = pike, family = shash, optimizer = "efs")
```

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

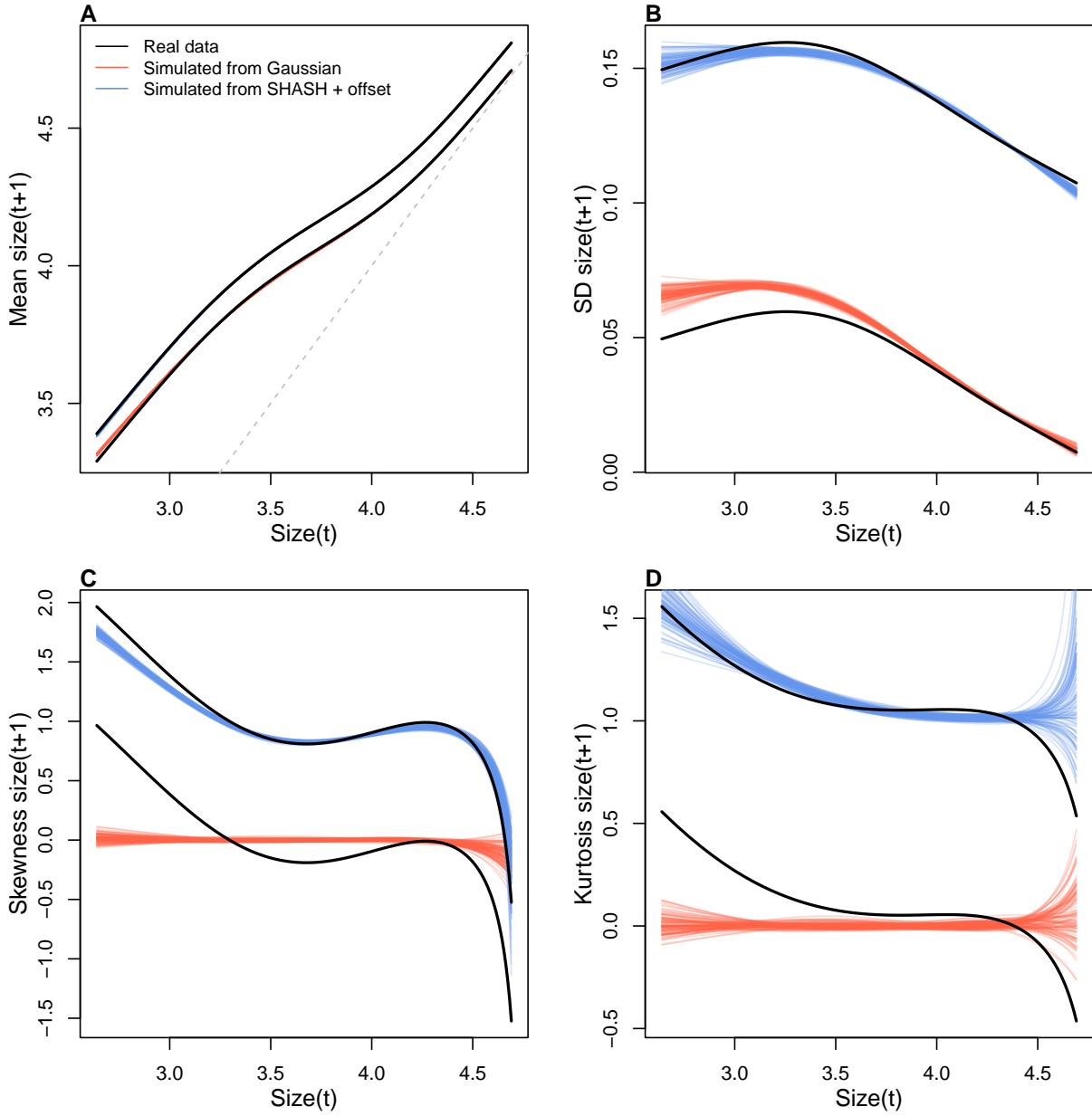


Figure 10: Comparisons between real pike data and data simulated from Gaussian and SHASH growth models for nonparametric measures of mean, standard deviation, skewness, and excess kurtosis of future size conditional on current size. Moments of the future size distribution are plotted with respect to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

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

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

943 4 Additional Figures

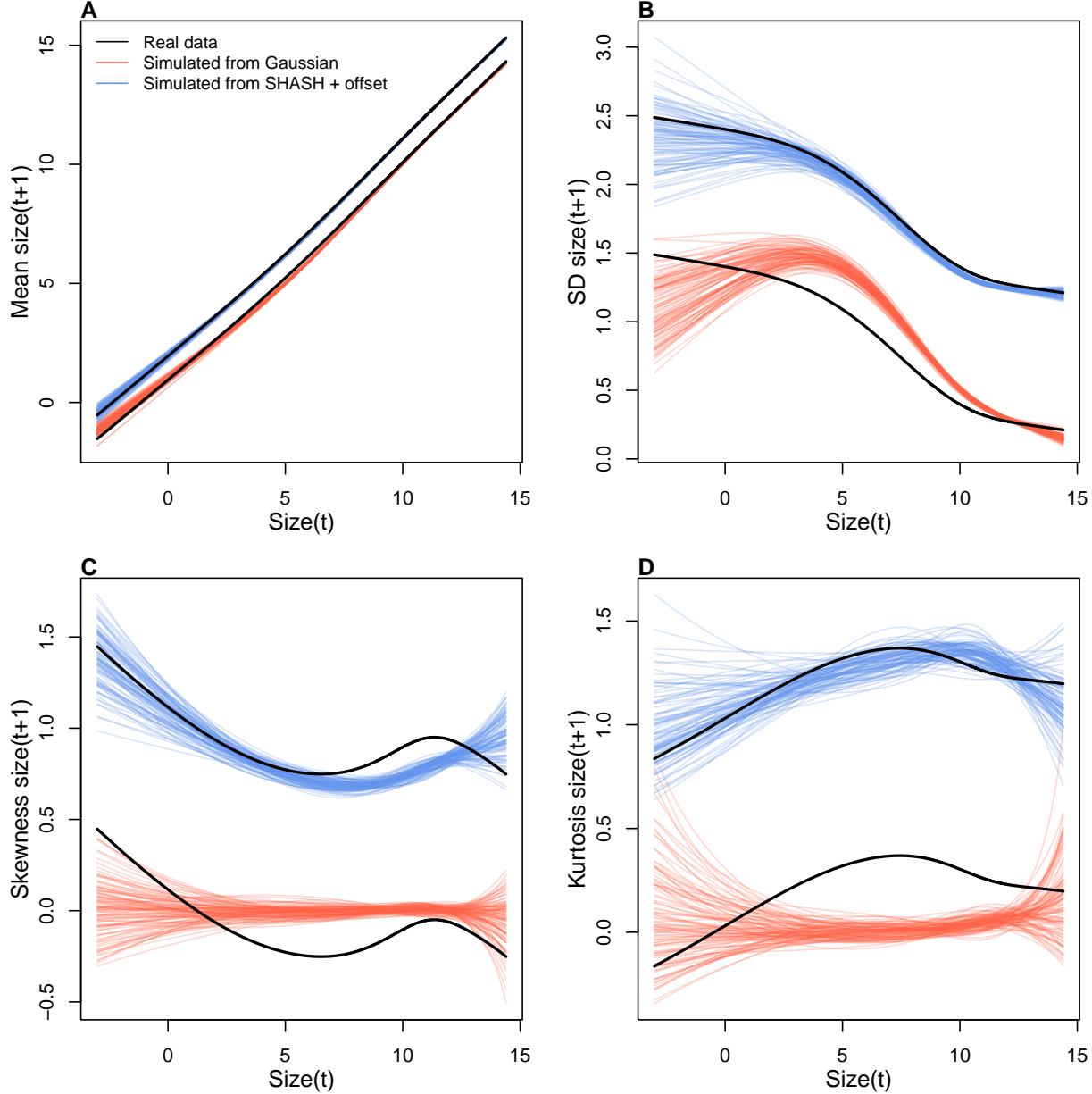


Figure 11: Comparisons among real cactus data and data simulated from Gaussian and SHASH growth models for mean, standard deviation, NP skewness, and NP excess kurtosis of future size conditional on current size. Figure made by script `cactus_growth_modeling_qgam.R`.

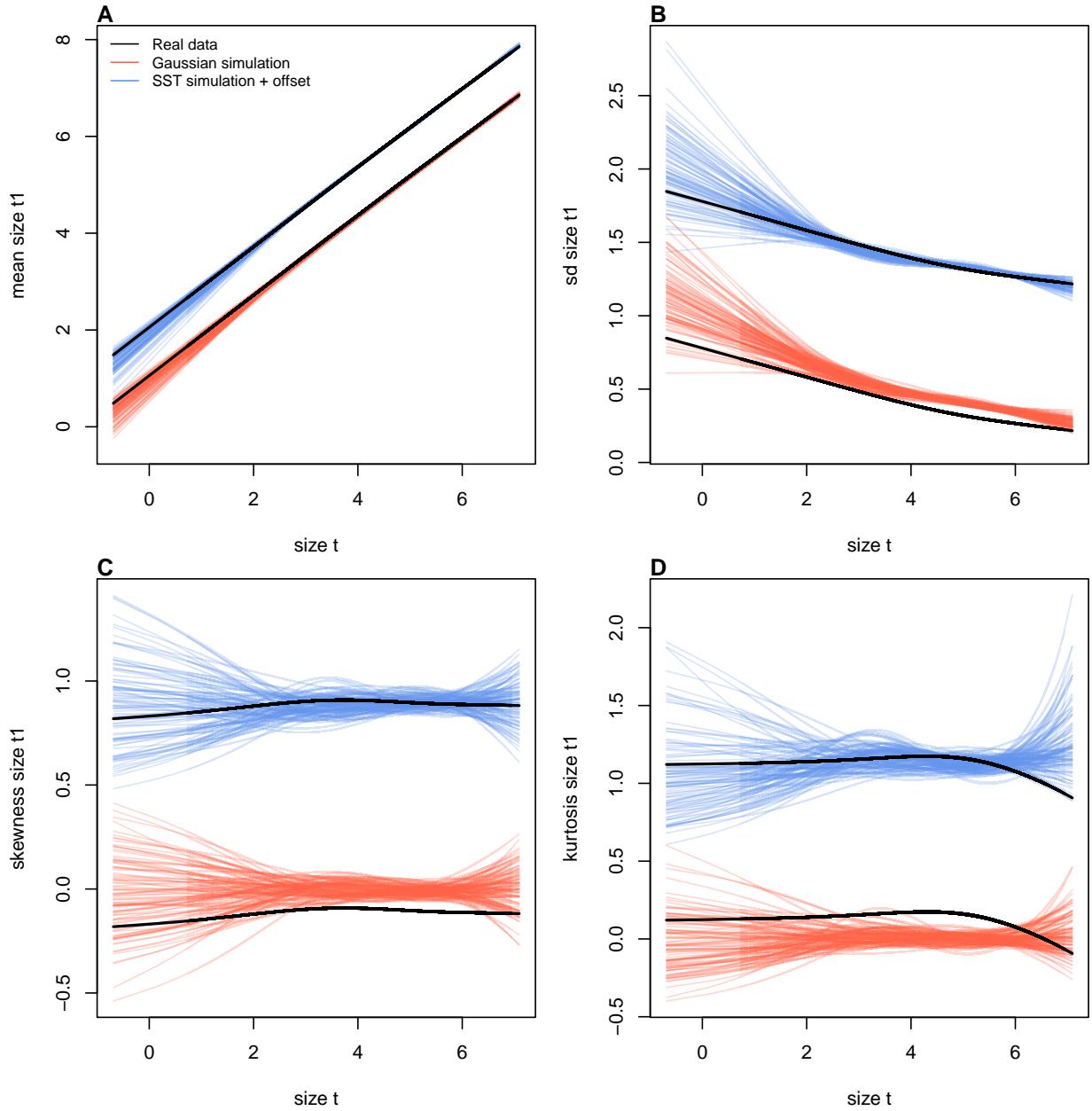


Figure 12: Comparisons between real orchid data and data simulated from Gaussian and skewed t growth models for mean, standard deviation, NP skewness, and NP kurtosis of future size conditional on current size. Top row (**A-D**) shows plants that were vegetative at the start of the transition year and 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`.

944 **Literature Cited**

- 945 Bruno, J.F., Ellner, S.P., Vu, I., Kim, K. & Harvell, C.D. (2011) Impacts of aspergillosis on sea fan coral
946 demography: modeling a moving target. *Ecological Monographs* **81**, 123–139.
- 947 Drees, T., Ochocki, B.M., Collins, S.L. & Miller, T.E. (2023) Demography and dispersal at a grass-shrub eco-
948 tone: a spatial integral projection model for woody plant encroachment. *Ecological Monographs* p. e1574.
- 949 Gould, W.R. & Nichols, J.D. (1998) Estimation of temporal variability of survival in animal populations.
950 *Ecology* **79**, 2531 – 2538.
- 951 Jones, M. & Pewsey, A. (2009) Sinh-arcsinh distributions. *Biometrika* **96**, 761 – 780.
- 952 Jones, M.C., Rosco, J.F. & Pewsey, A. (2011) Skewness-invariant measures of kurtosis. *The American
953 Statistician* **65**, 89 – 95.
- 954 Link, W.A. & Nichols, J.D. (1994) On the importance of sampling variance to investigations of temporal
955 variation in animal population size. *Oikos* **69**, 539 – 544.
- 956 Ochocki, B.M., Drees, T. & Miller, T.E. (2023) Density-dependent demography of creosote
957 bush (*larrea tridentata*) along grass-shrub ecotones. [https://doi.org/10.6073/pasta/
958 ca53c16f16dcf9fb11f3ee99ea5445ac](https://doi.org/10.6073/pasta/ca53c16f16dcf9fb11f3ee99ea5445ac).
- 959 Vindenes, Y., Edeline, E., Ohlberger, J., Langangen, Ø., Winfield, I.J., Stenseth, N.C. & Vøllestad, L.A.
960 (2014) Effects of climate change on trait-based dynamics of a top predator in freshwater ecosystems.
961 *The American Naturalist* **183**, 243–256.
- 962 Winfield, I., Fletcher, J. & James, J. (2013a) Pike fecundity data 1963-2002. NERC Environmental Infor-
963 mation Data Centre, <https://doi.org/10.5285/b8886915-14cb-44df-86fa-7ab718acf49a>.
- 964 Winfield, I., Fletcher, J. & James, J. (2013b) Pike growth data 1944-1995. NERC Environmental Information
965 Data Centre, <https://doi.org/10.5285/637d60d6-1571-49af-93f7-24c1279d884d>.
- 966 Winfield, I., Fletcher, J. & James, J. (2013c) Pike survival data 1953-1990. NERC Environmental Information
967 Data Centre, <https://doi.org/10.5285/813e07dd-2135-49bc-93c6-83999e442b36>.