Ecology, 99(10), 2018, pp. 2308–2317

Measurement error of state variables creates substantial bias in results of demographic population models

Allison Louthan[1,2,3] and Daniel Doak[2]

[1]*Department of Biology, Duke University, Durham, North Carolina 27708 USA*
[2]*Environmental Studies Program, University of Colorado–Boulder, Boulder, Colorado 80301 USA*

Abstract. Integral projection and matrix population models are commonly used in ecological and conservation studies to assess the health and extinction risk of populations. These models use one (or more) measurable state variable(s), such as size or age, to predict individual performance, which, ideally, is the sole determinant of an individual's expected fate. However, even if ecologists successfully identify and measure the observable state variable(s) that best predicts individual fate, we are rarely, if ever, able to perfectly measure state for many species, especially those with size structure, where total plant biomass or starch stores, for example, may be the best predictors of fate. Here, we used a series of simulations to test how this imperfect quantification of actual state ("measurement error") leads to inaccurate prediction of state-dependent fates and influences the predictions of structured population models. We simulated 10 yr of best practice field data collection using known vital rate functions and incorporated measurement error of different magnitudes and types (completely random, temporal, and individual based) for two size-structured life histories. We found that even for conservative error rates, most types of measurement error increased the median predicted population growth rate by 1–2% growth per year. However, the magnitude of this error differed substantially with life history strategy and error type, with some scenarios resulting in >8% median overestimation of population growth rate. This effect arises largely from the well-known econometrics problem of "regression dilution" (overestimation of the intercept and underestimation of the slope of a regression when the predictor variable is measured with error), which in our simulations typically results in overly optimistic predictions of small or young individuals' vital rates. Our results suggest that the problem of measurement error for state variables, present in many demographic studies but virtually unacknowledged in the ecological literature, may lead to substantial misestimation of population behavior, resulting in erroneous inferences about not only growth, but also extinction risk and other aspects of population dynamics.

Key words: demography; integral projection model; matrix model; measurement error; population growth; regression dilution.

Introduction

Demographic matrix models and integral projection models (IPMs) are commonly used in ecological and conservation studies to assess the health, viability, and extinction risk of populations. These models use one or more state variables to predict individual traits and performance. Ecologists commonly use age or size as state variables, as well as other traits such as breeding status, developmental stage, and location within different habitats or subpopulations. Identification of an appropriate state variable, or variables, is a critical consideration for all demographic models for any species (Pfister and Stevens 2003, Ellner and Rees 2006, Palmer et al. 2010). Ideally, the state variable(s) used in a model is the sole determinant of expected individual fates, such as probability of survival, mean growth increment, and reproductive output. However, in practice, state variables employed in population models may not have strong predictive power, in part because the choice of state variables is strongly driven by practicality and convention. It is fairly rare for researchers to test the power of alternate state variables to predict demographic traits, and even if ecologists successfully identify and measure the observable state variable that best predicts individual fate, we are very rarely able to perfectly measure state. We test how this imperfect quantification of state (hereafter, "measurement error") will influence the accuracy of predictions made from demographic models that rely on state-dependent vital rate functions, including both IPMs and population matrix models.

While several related topics have received significant attention in the ecological literature (Fig. 1), we have little understanding of how simple measurement errors will influence demographic predictions. To illustrate the ubiquity of this problem, consider the monitoring and then the construction of a demographic model for a perennial geophyte. For this species, starch stores in corms are likely the best predictor of individual fate. However, starch stores, or even corm mass, can only be measured destructively; therefore, a researcher instead measures the number and length of the leaves of each plant, combining these into a single estimate of leaf area, which correlates well, but imperfectly, with corm mass. In addition to this random error introduced into each measurement of each plant's size, plants could be systematically mismeasured in one of two ways: (1) on an individual basis, such that corm mass correlates with leaf area more or less strongly in some individuals, due to either differences in individual quality or microhabitat-driven differences in resource allocation, or (2) on a yearly basis, such that corm mass correlates with leaf area more strongly in

Manuscript received 26 September 2017; revised 5 June 2018; accepted 24 June 2018. Corresponding Editor: Viviana Ruiz-Gutierrez.
[3] E-mail: allisonmlouthan@gmail.com

one year than another, due to differences in within-plant resource allocation as a function of environmental forcing, or due to measurement by different researchers in different years. Exactly the same problem applies to many species. For example, in trees, diameter at breast height is often used to quantify size, and in reptiles and amphibians, snout–vent length is often used to quantify size, but these measurements are not a perfect metric of size or state. Similar mismeasurement occurs for many taxa with varying morphologies, where we can rarely measure the state variable that best predicts fate, and in many cases, do not know what this state variable might be.

While measurement error has received little attention in the literature, past work on related issues suggests that it may have important consequences for population projections (Fig. 1). For example, parameterization of models from small data samples can lead to misestimation and often bias of both deterministic and stochastic lambda ("sampling variability"; Brault and Caswell 1993, Doak et al. 2005, Fiske et al. 2008). Other issues with the overly simplified use of state variables have also been shown to substantially alter estimated population growth rates. For example, including individual-level heterogeneity in structured population growth models (some individuals consistently exhibit a stronger correlation between measured size and true state; e.g., Fox et al. 2006, or "growth autocorrelation" sensu Fujiwara et al. 2004) generally increases projected population growth rate (Conner and White 1999) and can increase variability in size structures through time (Pfister and Stevens 2002). The most likely route for measurement error to influence demographic predictions is through the well-known statistical problem of regression dilution, in which the slope of a linear regression is consistently underestimated when there is measurement error in the predictor variable. In the face of measurement error, this phenomenon will lead to misestimation of stage-variable vital rate relationships, which in turn may lead to significant bias in the projections of population growth rates and other model outputs. The direction of vital rate estimation bias is straightforward in linear models, but more difficult to predict in non-linear models, such as binomial survival probabilities (Hausman 2001). Further, almost all methods of correcting for this bias require independent measurements of the magnitude and structure of error (Carroll et al. 2006). In spite of these reasons to suspect that error in quantifying state variables could have strong impacts on demographic predictions, we are aware of no work that directly explores this issue. The closest analysis of which we are aware is that of Janeiro et al. (2017), who show that the statistical process of regression to the mean, resulting from autocorrelation in size in repeated measurements of growth, results in misestimation of vital rates consistent with regression dilution. This misestimation has the potential to misrepresent variance in individual growth trajectories across ages as well as heritability of life history traits, potentially leading to mischaracterization of any autocorrelated process, such as growth rates or trait evolution, by typically parameterized IPMs. However, as we have noted, the consequences of state variable measurement errors go beyond just misestimation of growth resulting from autocorrelated size measurements made with error.

To test how the type and magnitude of measurement error influence demographic model outputs, we conducted simulations that replicate the most common approach to constructing size-based demographic models, as well as exemplary data collection techniques, for two disparate life histories. We simulate annual monitoring of individual plant size, survival, and reproduction, but with the inclusion of

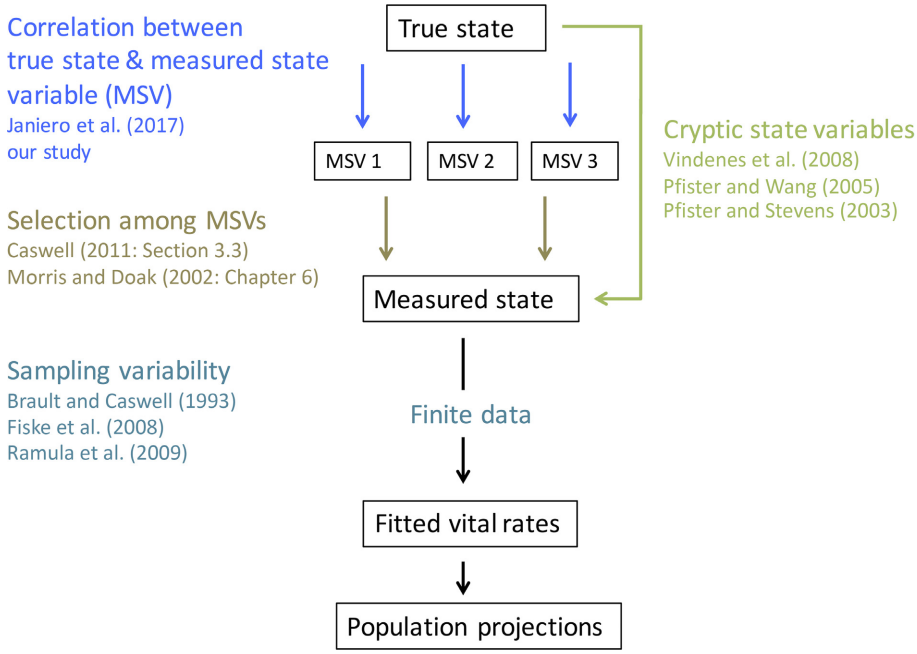

FIG. 1. Schematic of construction of population demographic models, showing the relationship between true state and measured state variables (MSV), fitting of vital rate functions (either continuous or binned), and subsequent construction of population projections. Colors correspond to bodies of literature addressing the role(s) of error in the construction of population matrix models and integral projection models.

differing degrees and types of error in the measurement of the underlying state variable "true size." We then construct IPMs with these data and ask how well the models predict true annual population growth rate for each species. We hypothesized that error would increase projected population growth rate for all types of error (e.g., Conner and White 1999) but that the magnitude of this misestimation would depend on life history due to differences in the sensitivity of population growth rate to different vital rates.

Methods

The true demographic process model

For our first species (hereafter, "short-lived species"), we simulated data sets based on Easterling et al.'s (2000) published relationships of survival, growth, variance in growth, and number of offspring as a function of size (stem diameter) for *Aconitum noveboracense*, a flowering herb. We simplified reproduction to only include production of (vegetative) shoots, with no fission of larger individuals (see Easterling et al. [2000] for details; shoot production is a common mode of reproduction in this species). We also assumed that they observed offspring number in time $t + 1$ rather than t and increased the intercept in the reproduction equation from 0.034 to 0.054 to obtain a deterministic population growth rate (λ) of ~1. For our second species (hereafter, "long-lived species"), we modified parameter estimates to increase growth rates and large individuals' survival and reproduction. Finally, we added a probability of reproduction vital rate, while also maintaining λ approximately equal to one, thereby generating higher sensitivity to large individuals' vital rates (Fig. 2). Appendix S1: Table S1 lists all these "true" demographic parameters and relationships for both life histories, as well as λ calculated using true size (λ_{true}). While we focus on two specific life histories, our results apply to any species whose measured state correlates imperfectly with an unmeasured state that best predicts fate.

We first simulated the true demographic process for each species, which we assume is deterministic, overlaying on top of this process a monitoring program that results in data taken with imperfect size measurements. We initiated our projections with a size distribution reflective of the stable size distribution (SSD), divided into 100 mesh points. For the short-lived species, we truncated our initial size distribution at the largest mesh point with an abundance ≥1% that of the second mesh point. Thus, initial sizes in the IPM models range from 0.04 to 5.08, with an upper size limit of 8. For the

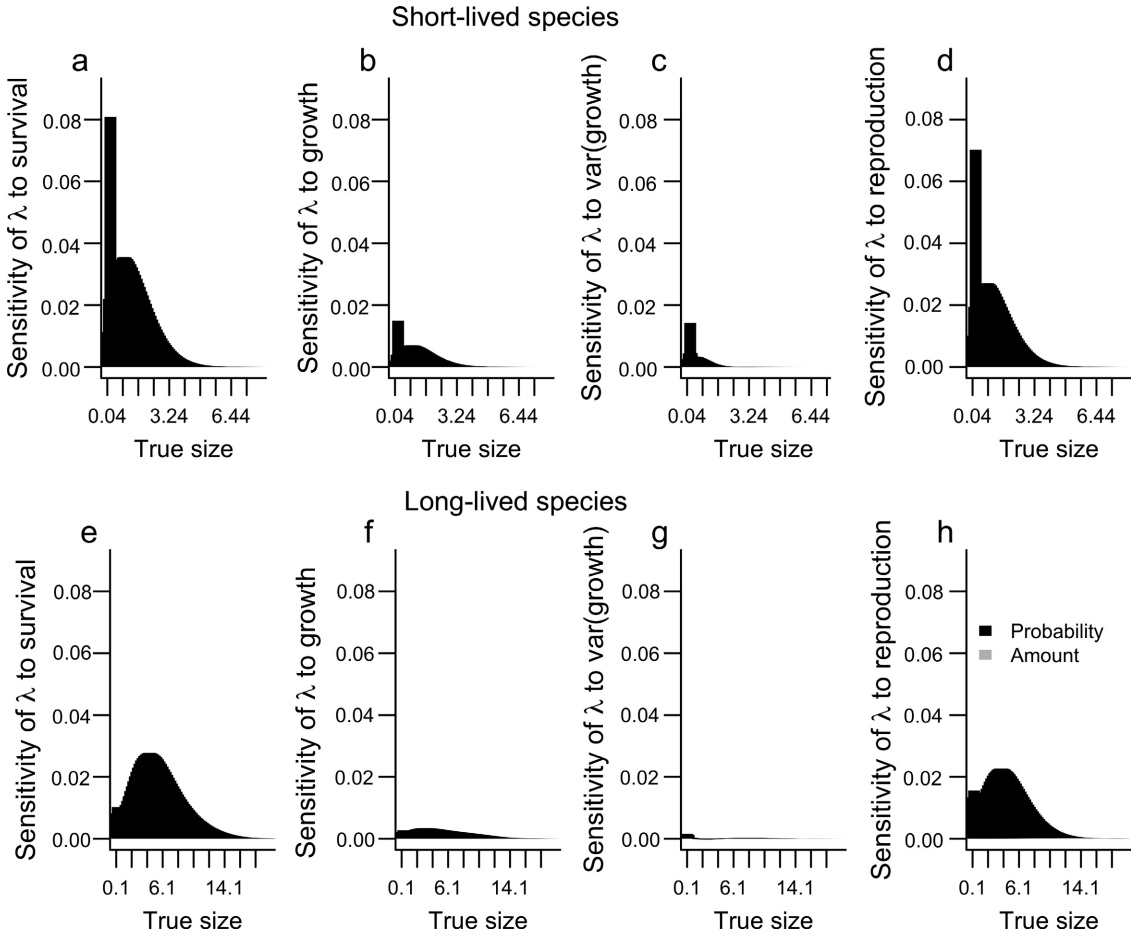

Fig. 2. Sensitivities of λ_{true} to size-specific vital rate values for both (a–d) short-lived and (e–h) long-lived species, estimated through perturbation from the matrix constructed used true vital rates. See Appendix S1: Fig. S11 for elasticities. For the long-lived species, note that there is very low sensitivity to amount of reproduction, so the gray is barely visible.

long-lived species, we used an initial size distribution with all mesh points that had an abundance ≥4% that of the second mesh point; thus, initial sizes ranged from 0.1 to 12.9, with 100 mesh points and with an upper size limit of 20.

We used these rules to simulate the fates of replicate samples of monitored individuals for 10 yr. We initiated each replicate with a population of 2,032 individuals (2,036 for long-lived), with sizes reflective of the population SSD as described above. Using the true size dependency of vital rates (Table S1), we simulated individual fates (using true sizes) for 10 yr, adding in new recruits each year to balance the number of non-surviving plants; sizes of new individuals were drawn from a uniform distribution bounded by [0.15, 0.25], following Easterling et al. (2000). These procedures correspond to a common field approach to demographic data collection: sample all individuals within a plot or set of plots and then search for and add new recruits to the set of monitored plants each year.

Simulation of monitoring with measurement error

Using these sets of true sizes and fates, we added different types and degrees of size measurement errors. We simulated seven different types of measurement error: σ^2_ε (each datum, or individual × year combination, has an associated, independent error value), σ^2_i (each year has an associated error value, applied to all individuals in that year), and σ^2_i (each individual has an associated error value, which is constant across all years of its life), and all combinations of these ($\sigma^2_{i+t}, \sigma^2_{\varepsilon+i}, \sigma^2_{\varepsilon+t}, \sigma^2_{\varepsilon+i+t}$). Note that we assume reproduction and survival are measured without error.

We simulated five magnitudes of total measurement error, corresponding to ρs (correlations) between true size and size with error = 0.99, 0.95, 0.85, 0.75, and 0.65. For a given error level, we normalized single and combined types of error such that the total amount of variability in all types of error was the same (e.g., for each error level, $\sigma^2_{\text{total}} = \sigma^2_\varepsilon = \sigma^2_{\varepsilon+t} = \sigma^2_{\varepsilon+i+t}$; Appendix S1). For combined error types, we assumed equal variance of each error type, with the summed variance equal to the desired value.

Error was added symmetrically around the true size of monitored individuals. To prevent unrealistic measured sizes (negative or infinite), we used a stretched beta to simulate measurement error of measured size (s_m) around true size (s): $s_m = s + f_{\min} + (f_{\max} - f_{\min}) * Beta(\alpha, \beta)$. Here, $\alpha = \beta$, so that the mean of $Beta(\alpha, \beta) = 0.5$, and its variance is $\sigma^2_{total} * (1/(f_{max} - f_{min}))^2$, where $f_{\min}$ and $f_{\max}$ are the minimum and maximum error values (Morris and Doak 2002). Here, $f_{\min} = -f_{\max}$, such that s_m is symmetrically distributed around s, bounded by $[s + f_{\max},\ s - f_{\max}]$; for all simulations, $f_{\max} = 2$ (short-lived) or 5 (long-lived). For individuals smaller than $f_{\min}$, we set the minimum and maximum values used to estimate error to a plant's true size, such that the minimum estimated size was never <0. Results do not differ substantially if we simulate measurement error proportional to true size: $s_m = s * (f_{\max} - f_{\min}) * Beta(\alpha, \beta)$, where α and β are as above, but $f_{\max} = 1 = -f_{\min}$, such that size with measurement error is bounded by $[0,\ 2 \times s]$; Appendix S1: Fig. S1. To identify the σ^2_{total} necessary to generate the desired correlation value for each type of error for each species, we solved directly for a series of σ^2_{total} values (Appendix S1); size distributions used in making these calculations were reflective of SSD, as we have described.

Analysis of simulation data and population projections

We simulated 500 replicate 10-yr data sets for the short-lived species and 1,000 replicates for the long-lived species. To each replicate, we added error in 35 different ways (7 patterns of error × 5 magnitudes of error), to create separate observed data sets. For each year in each data set, we fit regressions of survival, growth, variance in growth, and amount of reproduction on measured size, all of the same functional form as the true relationships (Appendix S1: Table S1; for the long-lived species, we fit probability of reproduction as well, and amount of reproduction was conditional on reproducing). Thus, we assume that the correct functional forms of all vital rates are known, but that the parameter estimates are unknown, as is any annual variation in demography (which does not in fact exist in the actual demographic process). For our long-lived species, not all of the binomial regressions converged; for the regressions that did not converge, we refit them without a size term; these were a small percentage of the total binomial regressions (0.00002%). Using a binning approach to obtaining vital rates, rather than fitting continuous functions, does not change our results; thus, the results we outline apply also to discrete parameterizations of matrix models (Appendix S1: Fig. S2).

We constructed year-specific kernels with 100 mesh points using these model fits. As different error magnitudes and types lead to different maximum sizes of plants, the mesh points for each combination of error parameters were set to evenly divide sizes between a minimum of 0 and a maximum of 1.5 × the maximum measured size observed across all years in a simulation (results do not differ if we use a constant upper size class bound of 8 for the short-lived species; Appendix S1: Fig. S3). When constructing kernels, if probability density functions (pdfs) of growth fell outside of the size class bounds, we discarded all probability outside of the size class bounds and renormalized the pdf to one (Williams et al. 2012). Results do not differ if we remove one-half of the simulations in which >0.1 probability was outside either the lower or upper bound of the kernel in any year (83% of error type × ρ × simulation × year combinations; Appendix S1: Fig. S4).

We obtained a separate estimate of stochastic lambda (λ_s) for the model fit to each data set by selecting from each year-specific kernel with equal probability, using the last 8,000 years of 10,000-year projections to guard against transient dynamics. We initiated these projections with the stable size distribution obtained from across-year average kernels, obtained for each fitted model. To visualize the effect of error on deviation from λ_{true}, we compared the across-replicate average λ_s to λ_{true} (which is based on unvarying relationships of survival, growth, reproduction, and true size), summarizing the direction and magnitude of error in estimates of λ_s for each error type and magnitude.

For each data set, in addition to estimating λ_s, we used two modified approaches to obtain estimates of (deterministic) λ. First, we calculated λ from the average of the year-specific kernels obtained above; second, we fit models for survival,

growth, variance in growth, probability of reproduction (long-lived only), and amount of reproduction (conditional on reproducing for the long-lived species) with data from all years and then constructed kernels and obtained deterministic λ using these relationships. Results using the across-replicate average of both types of deterministic λ do not differ from results using λ_s (Appendix S1: Fig. S5, S6). λ_s and λ of the mean kernel across years are highly correlated (Appendix S1: Table S2), suggesting the majority of the misestimation of λ_s with measurement error arises due to misestimation of the mean vital rates, rather than misestimation of the variance or covariance of these rates.

To better understand how measurement errors drive effects in estimated λ_s, we conducted two sensitivity analyses. First, we calculated sensitivities and elasticities of λ_{true} to survival, growth, variance in growth, and amount of reproduction (and probability of reproduction for the long-lived species) via simulation, perturbing mesh-point-specific vital rates 5% higher and lower than their true value, separately for each vital rate × mesh point combination. Second, for each life history, we quantified the contribution of each parameter types' misestimation to the total misestimation of λ_s across all error type × ρ combinations. We first substituted the parameter estimates for all vital rates from the model run associated with the median estimated λ_s for each error type × ρ combination into our "true" kernel, calculating an expected percent deviation from λ_{true} given the misestimation in all vital rates. We then calculated a deviation from λ_{true} separately for each vital rate, using the misestimated parameter values only for the vital rate of interest, with misestimated parameters of growth/variance in growth and reproduction/probability of reproduction substituted at the same time. We then regressed the total expected percent deviation (λ calculated using all misestimated vital rate values) on the contribution of each vital rate's separate predicted misestimation values, and interactions among these effects, using type II sum of squares to partition the variance attributable to each of these effects. All analyses were conducted with R version 3.3.2 (R Core Team 2016) and R code is available in Data S1.

Results

All types of measurement error result in overestimation of mean λ_s in our short-lived species (Fig. 3); overestimation of λ_s ranges up to 11.24% for a single simulated data set, and up to 3.32% median error for σ_t^2, temporally correlated error, at $\rho = 0.65$. The simplest type of error, completely random (σ_ε^2,), shows median overestimation of λ_{true} of 1.86% when $\rho = 0.65$. Other error types also show substantial overestimation of λ_{true} when $\rho \leq 0.75$. Higher error also generally increases the variability in predictions of λ_s. σ_t^2, temporally correlated error, creates far higher median error and variability in estimated λ_s compared to other error types, likely due to a limited number of randomly sampled error magnitudes for years compared to individuals or individual × years.

Effects of measurement error for the long-lived species are similar to those for the short-lived species, but with some differences (Fig. 3). The λ_s values are again overestimated, as with the short-lived species. However, error types that contain σ_t^2 show even stronger overestimation than for the short-lived species (maximum median overestimation for σ_t^2 in long-lived, 8.30%; short-lived, 3.32%; see Appendix S1: Fig. S7), and greater variance, with some λ_s estimates actually lower than λ_{true} rather than higher (Fig. 3).

Deviation from λ_{true} is due to a misestimation of vital rate relationships when including measurement error. In the short-lived species, for survival, growth, and reproduction, increasing amounts of error generally lead to an overestimation of the intercept (b_0) and an underestimation of the slope (b_1, Fig. 4; Appendix S1: Fig. S8). Conversely, for regressions of variance in growth, increasing amounts of error lead to an overestimation of both the intercept and the slope, likely due to the higher overall variability in observed growth when including measurement error of any kind. Misestimation of survival, growth, and reproduction parameter estimates for σ_t^2 overestimate both intercept and slope parameters, resulting in population projections that strongly deviate from λ_{true} (Fig. 3).

For the long-lived species, some vital rates that strongly impact λ_s (survival, probability of reproduction; Fig. 2) do not show the classic misestimation patterns expected from regression dilution for any error type (Fig. 5). In particular, both survival parameters are strongly overestimated for all types of error, and misestimation of probability of reproduction parameters varies with error type. Classic regression dilution (an overestimation of b_0 and underestimation of b_1) only occurs for most error types for growth and amount of reproduction vital rates.

Our analyses of each vital rate's contributions to expected percent deviation from λ_{true} suggest that the misestimation of growth parameters and, much less importantly, survival parameters, drives much of the pattern in misestimation of λ_{true} in the short-lived species (Table 1). Overestimation arises consistently for the short-lived species because it has higher sensitivity to growth and survival in smaller size classes than larger size classes (Fig. 2), and growth and survival both show classic regression dilution (Fig. 4). Thus, small individuals' vital rates are overestimated, resulting in overestimation of λ_{true}.

By contrast, misestimation of reproduction, survival, and growth, as well as interactions among vital rate misestimations, all contribute to the overall misestimation of λ_{true} in the long-lived species (Table 1). Classic regression dilution in amount of reproduction parameters and growth parameters, as well as overestimation of both survival parameters (Fig. 5), likely lead to an overall increase in estimated λ_s, countering any effects of pessimistic probability of reproduction parameter estimates. Interactions between survival and reproduction vital rates (both probability and amount) likely increase estimated λ_s because with increases in survival, individuals can more quickly transition to the high-reproducing larger size classes.

Discussion

For both species, our results suggest that measurement error creates considerable variability in estimated λ_s and also results in its consistent overestimation (Fig. 3). In the short-lived species, the median overestimation of λ_{true} arises from overly optimistic estimates of small individuals' vital rates

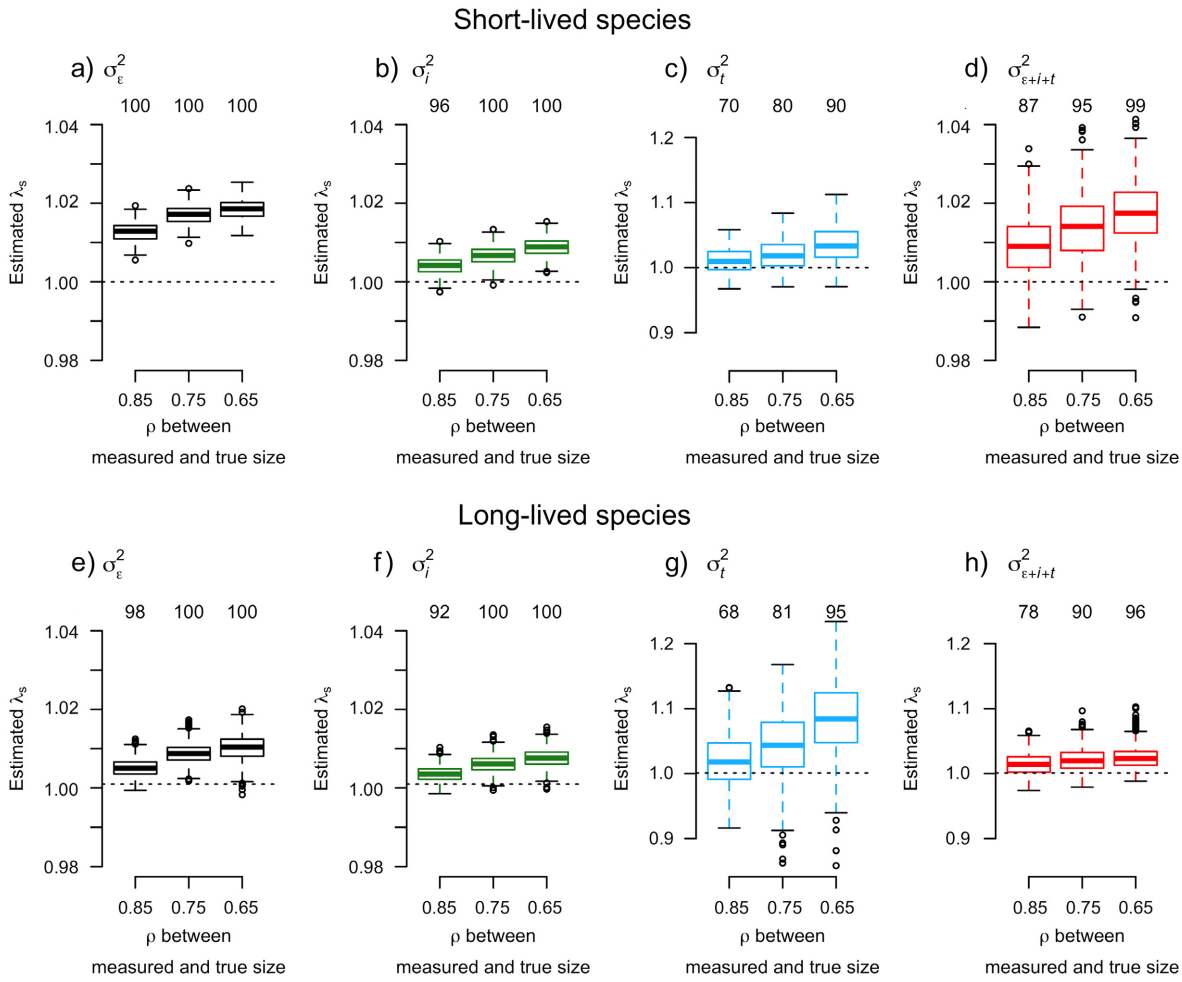

FIG. 3. Deviation from λ_{true} for simulations with the correlation between true and measured size (ρ) ranging from 0.85 to 0.65 and different types of error structures for (a–d) short-lived species and (e–h) long-lived species. The variable σ^2_ε represents random individual measurement errors, σ^2_t temporally correlated error, σ^2_i individually correlated error, and $\sigma^2_{\varepsilon+i+t}$ a combination of all three error types. We show standard boxplots, such that the midline represents the median, box edges represent the 25th and 75th percentiles, and whiskers represent the reasonable extremes of the data. Data points larger or smaller than the reasonable extremes are shown by points. Dotted line shows the median value of λ_{true}; numbers indicate the percentage of replications where estimated $\lambda_s > \lambda_{true}$. Colors correspond to error structures and are consistent with subsequent figures. Note change in scale in c, g, and h; see Appendix S1: Figs. S12, S7 for all ρ values and error types.

(Fig. 4) coupled with high sensitivity to these same rates (Fig. 2). Vital rates of large individuals have low sensitivity in this species (Fig. 2), such that pessimistic estimates of their vital rates do not counterbalance the effects on vital rates of small individuals. For the long-lived species, we do not see classic regression dilution for all vital rates, making the explanation of the consistent overestimation of λ_{true} more complex. In this life history, there are relatively high sensitivities to large individuals' vital rates (Fig. 2), but these rates are not consistently overestimated. For example, both survival parameters are overestimated, so all survival rates are estimated to be higher than is correct, but we see regression dilution for growth parameters, meaning that large individuals' growth rates are underestimated. Results on both species are consistent with a series of studies (Doak et al. 2005, Fiske et al. 2008) finding overestimation of population growth rate with small sample sizes that also results (for different reasons) from misestimation of vital rates. Our findings may provide some explanation for overly optimistic predictions of population growth rate seen empirically in a variety of species (e.g., *Arisaema triphyllum* [Bierzychudek 1999] and *Lepanthes rubripetala* [Schödelbauerová et al. 2010], but see Brook et al. 2000, Crone et al. 2013).

Our results suggest that overestimation of population growth is a common phenomenon, but depends on the presence and strength of classic regression dilution as well as less predictable patterns of misestimation due to measurement errors. Previous studies show that, while regression dilution is predictable for linear vital rates (as we usually see here), it is inconsistent for non-linear vital rates (as we see here for binomial outcomes in our long-lived species). For situations similar to our completely random (ε) error structure, past simulations for logistic regressions sometimes show classic regression dilution (Reeves et al. 1998, Kim et al. 2006). However, under other circumstances, including with large sample sizes, the direction of bias can reverse (Kim et al. 2006). Unequal numbers of observations across the range of predictor values (Kim et al. 2006) and heteroscedastic

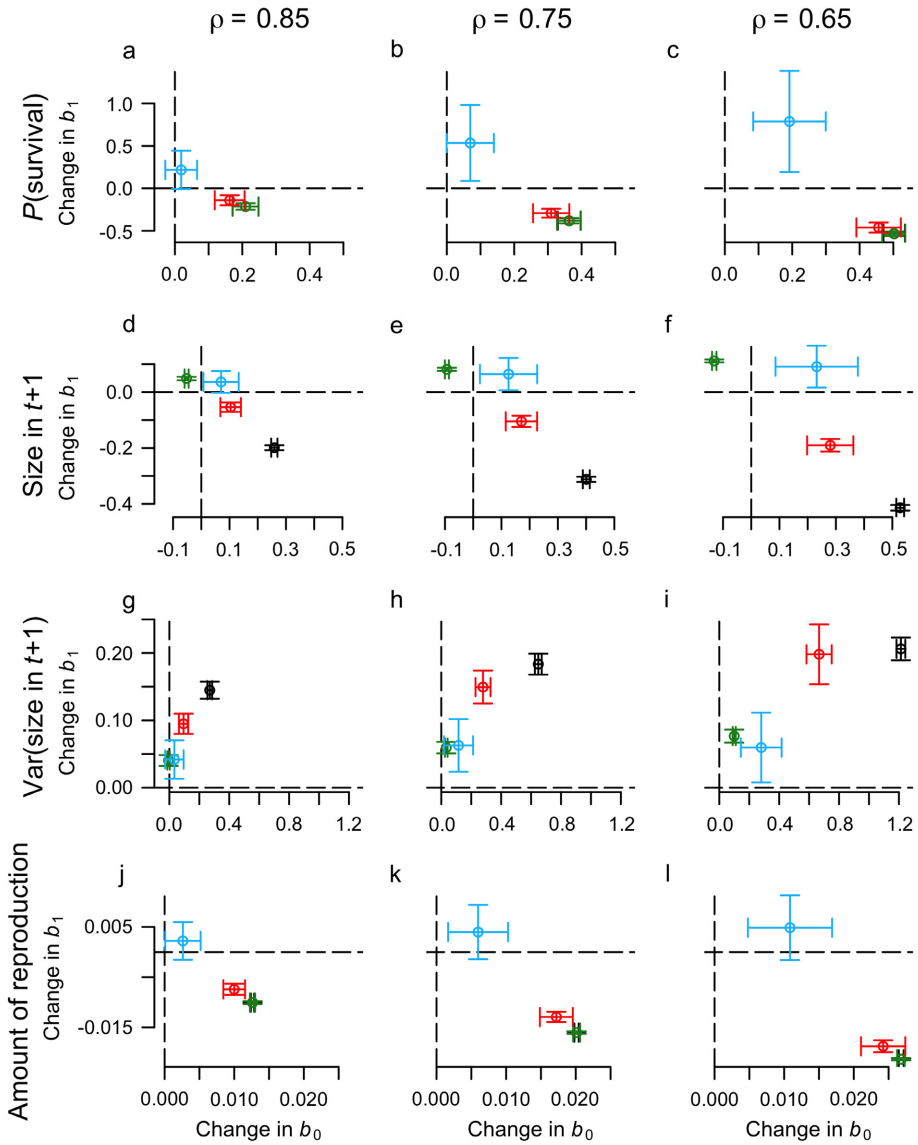

FIG. 4. Misestimation of the vital rate function parameters in the short-lived species for ρ (correlation) = 0.85, 0.75, and 0.65 for survival (a–c), growth (d–f), variance in growth (g–i), and amount of reproduction (j–l); points represent mean misestimation and bars indicate standard deviation. For all vital rates, we show the mean misestimation of intercept, b_0, and coefficient of size, b_1, by subtracting these numbers from their true value. Dotted lines show 0, representing no change in either b_0 or b_1. Colors correspond to those used in Fig. 3 to represent different error types; recall that σ^2_ε represents random error, σ^2_t temporally correlated error, and σ^2_i individually correlated error, and $\sigma^2_{\varepsilon+i+t}$ a combination of these; when σ^2_ε (black) is not visible, these data are simply obscured by the σ^2_i (green) data. For an alternate representation that shows the predicted values, see Appendix S1: Fig. S13, and for all ρ values and error types, see Appendix S1: Fig. S14. For a discussion of how error in independent and dependent variables are expected to change the fits of linear and binomial regressions, see Appendix S1.

between-plant variance in size (Burstyn et al. 2006) can also substantially influence the strength and direction of bias. The magnitude of classical regression dilution in logistic regression, and when logistic regression should experience classical regression dilution vs. the opposite (here, in the long-lived species, overestimation of the size coefficient), is an active area of research in biomedical fields (Burstyn et al. 2006, Kim et al. 2006).

While generalities about regression dilution may be difficult, ecologists could easily simulate regression results with measurement errors and their actual distribution of sizes (or other state variable) to test for the direction of parameter misestimation for each vital rate. For example, ecologists could (1) assume a size- (or other state variable-) dependent function for survival; (2) simulate survival probabilities of each of their measured individuals using this function; (3) fit a function for survival using these simulated probabilities; (4) add normally distributed error around measured sizes, then repeat steps 1–3 and assess the direction of parameter bias in the resulting functions. We recommend conducting this analysis separately for growth rates, where effects of measurement errors are compounded across years (Janeiro et al. 2017).

The magnitude of bias in λ_s we find in this study is similar to what we might expect to find in many demographic

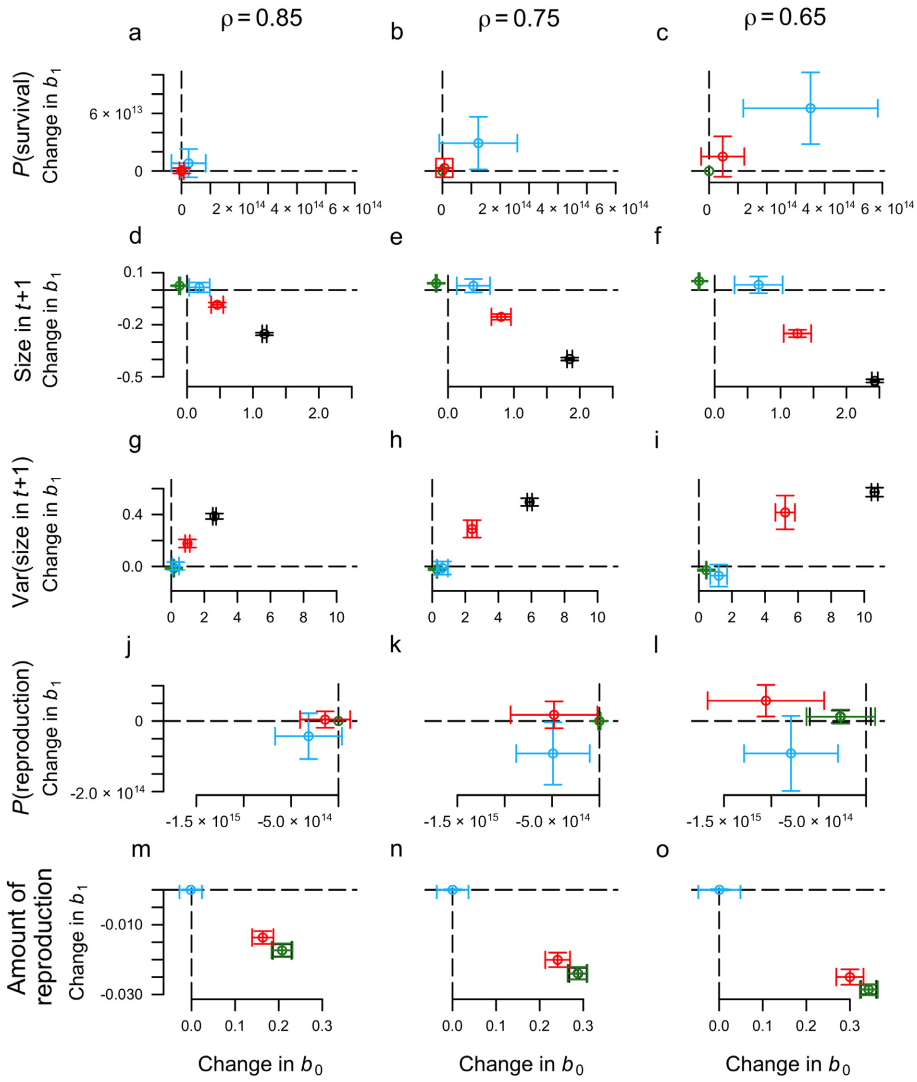

FIG. 5. Misestimation of the parameters in our long-lived species for probability of survival (a–c), growth (d–f), variance in growth (g–i), probability of reproduction (j–l), and amount of reproduction (m–o). For all vital rates, we show the mean misestimation of intercept b_0 and coefficient of size, b_1, by subtracting these numbers from their true value. Points represent mean misestimation and bars indicate standard deviation, and dotted lines represent no misestimation of parameters. When σ^2_ε (black) is not visible, these data are simply obscured by the σ^2_i (green) data. See Figure 4 for further explanation. For an alternate representation that shows predicted values, see Appendix S1: Fig. S15, and for all ρ values and error types, see Appendix S1: Fig. S16.

studies, including the best executed studies. Across the few studies we found that present multiple metrics of size, the mean Pearson's R^2 value between size measures is 0.80 (range 0.97–0.42), while the mean correlation coefficient between the best estimate of true size and other measures is 0.67 (range 0.95–0.31; Appendix S1: Table S3). We find substantial bias in λ_s estimates even though, in many ways, our study represents an ideal sampling protocol that is not attained by all or most field studies. In our short-lived species, for less well-designed sampling regimes (e.g., initiating our population at a uniform size distribution), measurement error can result in even larger overestimates of λ_{true} (Appendix S1: Fig. S9). Finally, studies conducted on non-stable populations might overestimate λ_{true} even more. For example, maximum median overestimation is even higher (3.76%) when $\lambda_{true} = 1.02$ in the short-lived species. For λ_{true} values <1, these effects depend on life history: maximum median estimation is increased to 3.58% for the short-lived but decreased to 2.65% for the long-lived species (with $\lambda_{true} = 0.98$ for short-lived, 0.99 for long-lived species). Importantly, other predictions of demographic models could also be affected by measurement error; for example, Janeiro et al. (2017) show that measurement error results in misestimation of the covariance among subsequent ages' measured phenotype, and results in poor predictions of phenotypic similarity across relatives, two phenomena that drastically reduce our ability to accurately predict evolutionary dynamics using IPMs (Janeiro et al. 2017).

As outlined in the introduction, individual variation in measurement error might arise if some individuals have high intrinsic quality (e.g., consistently high tuber storage throughout their life), whereas temporal variation in error could occur if annual fluctuations in the environment elicit differential allocation (e.g., a wet year promotes higher allocation to

TABLE 1. Variance explained by misestimation of each vital rates' parameters and the interactions among misestimation of parameters.

Vital rate	Variance explained (%)	
	Short-lived species	Long-lived species
Survival	11.1	16.8
Growth and var(growth)	86.7	20
Reproduction	<0.1	29.7
Survival × growth and var(growth)	<0.1	8.3
Survival × reproduction	<0.1	12.1
Growth × reproduction	<0.1	6.2
Survival × growth and var(growth) × reproduction	0.9	2.5
Residual variance	1.2	4.5

Note: For the short-lived species, "reproduction" includes only amount of reproduction, but for the long-lived species, "reproduction" includes both probability and amount of reproduction. Note the relatively large contributions of interactions among vital rates in the long-lived, but not short-lived species (Appendix S1: Fig. S10).

above- vs. belowground organs); regardless, all types of error result in overestimation of population growth rates. For different reasons, if temporal or individual variation in growth rates is present in a population but not included in population projections, omission of this individual-level heterogeneity can often result in underestimation of population growth rates (Conner and White 1999, Fox and Kendall 2002). Pfister and Wang (2005) show that underestimation of population growth rate can arise when large, fast-growing individuals' growth rates are not incorporated into population projections that exclude individual heterogeneity (though see Vindenes and Langangen (2015), who see dissimilar results for an IPM with a genetic component). Similar to our short-lived species, Kendall et al. (2011) found strong effects of misestimation of survival rates on population growth rate: simulating individual-level heterogeneity in survival increases estimated population growth rate, whereas simulating individual-level heterogeneity in reproductive rate has little effect on population growth rate.

In this study, we use the standard linear and non-linear models most commonly employed to fit vital rate functions, but a series of more complex, less commonly used alternative model-fitting metrics could reduce the magnitude of misestimation. Correcting for regression dilution is commonly studied in other disciplines. For example, Frost and Thompson (2000) and Longford (2001) show simple ways to correct for this bias assuming normally distributed variables, and these methods have been extended to non-linear models (Carroll et al. 2006). Another promising approach for reducing some measurement error bias is the use of multi-time-step information to gauge across-age covariation in vital rates to better estimate true state and covariation effects on vital rates; a random regression approach, which has been employed in multiple fields of biology including genetics and behavior (Wilson et al. 2005, Martin et al. 2010, Dingemanse and Dochtermann 2012), provides one avenue for such an analysis tactic (Janeiro et al. 2017). However, all of these corrections require information on the magnitude and structure of error (e.g., remeasured independent variables, which is only easy to obtain for σ^2_ε; but see Hong and Tamer 2003). State-space IPMs can also explicitly incorporate observation error (Dennis et al. 2006, White et al. 2016), but one must specify the correlation structure of this error. This structure is not always known, and at least for our long-lived species, the magnitude of error differs strongly with the structure of this error.

Hierarchal Bayesian approaches, while rarely employed in demographic studies to date, allow estimation of both process and measurement error, and most critically, can explicitly incorporate uncertainty in the magnitude and structure of error (Elderd and Miller 2016). Such approaches could allow one to parameterize regression models that explicitly incorporate measurement error, associated with individual, time, and/or space. One must still specify the form of the error (additive vs. multiplicative) and structure of error (completely random vs. temporally or individually correlated), but many aspects of the error structure can be represented as fitted parameters (e.g., variance of random error in measurements or variance of temporally correlated error), meaning that model outputs provide a way to test whether different types of error are small or nonexistent. However, it remains to be seen if with realistic sample sizes and measurement error processes, these models can accurately estimate and thus correct for the measurement error in the construction of demographic models.

More simply, an approach similar to that used on our simulation models could be used to gauge the likely seriousness of measurement errors for population predictions. First, using observed samples and vital rates, a researcher can generate simulations with varying degrees and types of measurement errors and ask how strong and consistent parameter bias is, particularly for non-linear vital rates (see Appendix S1: Fig. S8). Second, analysis of the sensitivity patterns of the general life history (see Fig. 2) will provide a sense of whether biased vital rate estimates will strongly overlap with those rates to which population growth rate is consistently sensitive. Third, a path analysis approach could provide an estimate of the magnitude of measurement error, provided some of the vital rates are linear and one has sufficient data to accurately estimate σ^2_i and σ^2_i (Janeiro et al. 2017). Perhaps most simply, ecologists should pay greater attention to testing multiple possible state variables and make efforts to directly test the strength and structure of correlations between observable and the most direct measures of state. For example, for a sessile marine invertebrate, one could regress non-destructively measured shell size with total biomass (which likely correlates more strongly with fate) in multiple years to assess the correlation coefficient between measured and "true" size. Finally, we note that our analyses have only focused on population growth rates, but that the many other predictions of demographic models are also likely to be altered by measurement error (Janeiro et al. 2017).

ACKNOWLEDGMENTS

This work was supported by NSF-DEB 1340024, 135378, 1242355 to D. Doak, NSF-DEB 1311394 to D. Doak and A. Louthan, and the UNESCO-L'OREAL International Fellowship for Young Women in Life Sciences to A. Louthan.

Literature Cited

Bierzychudek, P. 1999. Looking backwards: Assessing the projections of a transition matrix model. Ecological Applications 9:1278–1287.

Brault, S. and H. Caswell. 1993. Pod-specific demography of killer whales (*Orcinus orca*). Ecology 74:1444–1454.

Brook, B. W., J. J. O'Grady, A. P. Chapman, M. A. Burgman, H. R. Akçakaya, and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. Nature 404:385–387.

Burstyn, I., K. M. Kim, N. Cherry, and Y. Yasui. 2006. Metamodels of bias in Cox proportional-hazards and logistic regressions with heteroscedastic measurement error under group-level exposure assessment. Annals of Occupational Hygiene 50:271–279.

Carroll, R., D. Ruppert, L. Stefanski, and C. Crainiceanu. 2006. Measurement error in nonlinear models: a modern perspective. Second edition. Chapman & Hall, Boca Raton, Florida, USA.

Conner, M. M., and G. C. White. 1999. Effects of individual heterogeneity in estimating the persistence of small populations. Natural Resource Modeling 12:109–127.

Crone, E. E., et al. 2013. Ability of matrix models to explain the past and predict the future of plant populations. Conservation Biology 27:968–978.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Stapes. 2006. Estimating density dependence, process noise, and observation error. Ecological Monographs 76:323–341.

Dingemanse, N. J., and N. A. Dochtermann. 2012. Quantifying individual variation in behaviour: mixed-effect modelling approaches. Journal of Animal Ecology 82:39–54.

Doak, D. F., K. Gross, and W. F. Morris. 2005. Understanding and predicting the effects of sparse data on demographic analyses. Ecology 86:1154–1163.

Easterling, M. R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new structured population model. Ecology 81:694–708.

Elderd, B. D., and T. E. X. Miller. 2016. Quantifying demographic uncertainty: Bayesian methods for integral projection models. Ecological Monographs 86:125–144.

Ellner, S. P., and M. Rees. 2006. Integral projection models for species with complex demography. American Naturalist 167:410–428.

Fiske, I. J., E. M. Bruna, and B. M. Bolker. 2008. Effects of sample size on estimates of population growth rates calculated with matrix models. PLoS ONE 3:e3080.

Fox, G. A., and B. E. Kendall. 2002. Demographic stochasticity and the variance reduction effect. Ecology 83:1928–1934.

Fox, G. A., B. E. Kendall, J. W. Fitzpatrick, and G. E. Woolfenden. 2006. Consequences of heterogeneity in survival probability in a population of Florida scrub-jays. Journal of Animal Ecology 75:921–927.

Frost, C., and S. Thompson. 2000. Correcting for regression dilution bias: comparison of methods for a single predictor variable. Journal of the Royal Statistical Society Series A 163:173–190.

Fujiwara, M., B. E. Kendall, and R. M. Nisbet. 2004. Growth autocorrelation and animal size variation. Ecology Letters 7:106–113.

Hausman, J. 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. Journal of Economic Perspectives 15:57–67.

Hong, H., and E. Tamer. 2003. A simple estimator for nonlinear error in variable models. Journal of Econometrics 117:1–19.

Janeiro, M. J., D. W. Coltman, M. Festa-Bianchet, F. Pelletier, and M. B. Morrissey. 2017. Towards robust evolutionary inference with integral projection models. Journal of Evolutionary Biology 30:270–288.

Kendall, B. E., G. A. Fox, M. Fujiwara, and T. M. Nogeire. 2011. Demographic heterogeneity, cohort selection, and population growth. Ecology 92:1985–1993.

Kim, H. M., Y. Yasui, and I. Burstyn. 2006. Attenuation in risk estimates in logistic and Cox proportional-hazards models due to group-based exposure assessment strategy. Annals of Occupational Hygiene. 50:623–635.

Longford, N. T. 2001. Correspondence. Journal of the Royal Statistical Society, Series A 164:565.

Martin, J. G. A., D. H. Nussey, A. J. Wilson, and D. Réale. 2010. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. Methods in Ecology and Evolution 2:362–374.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer Associates, Sunderland, Massachusetts, USA.

Palmer, T. M., D. F. Doak, M. L. Stanton, J. L. Bronstein, E. T. Kiers, T. P. Young, J. R. Goheen, and R. M. Pringle. 2010. Synergy of multiple partners, including freeloaders, increases host fitness in a multispecies mutualism. Proceedings of the National Academy of Sciences USA 107:17234–17239.

Pfister, C. A., and F. R. Stevens. 2002. The genesis of size variability in plants and animals. Ecology 83:59–72.

Pfister, C. A., and F. R. Stevens. 2003. Individual variation and environmental stochasticity: implications for matrix model predictions. Ecology 84:496–510.

Pfister, C. A., and M. Wang. 2005. Beyond size: matrix projection models for populations where size is an incomplete descriptor. Ecology 86:2673–2683.

R Core Team. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Reeves, G. K., D. R. Cox, S. C. Darby, and E. Whitley. 1998. Some aspects of measurement error in explanatory variables for continuous and binary regression models. Statistics in Medicine 17:2157–2177.

Schödelbauerová, I., R. L. Tremblay, and P. Kindlmann. 2010. Prediction vs. reality: Can a PVA model predict population persistence 13 years later? Biodiversity and Conservation 19:637–650.

Vindenes, Y., and E. Langangen. 2015. Individual heterogeneity in life histories and eco-evolutionary dynamics. Ecology Letters 18:417–432.

White, J. W., K. J. Nickols, D. Malone, M. H. Carr, R. M. Starr, F. Cordoleani, M. L. Basket, A. Hastings, and L. W. Botsford. 2016. Fitting state-space integral projection models to size-structured time series data to estimate unknown parameters. Ecological Applications 26:2677–2694.

Williams, J. L., T. E. X. Miller, and S. P. Ellner. 2012. Avoiding unintentional eviction from integral projection models. Ecology 93:2008–2014.

Wilson, A. J., L. E. B. Kruuk, and D. W. Coltmann. 2005. Ontogenetic patterns in heritable variation for body size: using random regression models in a wild ungulate population. American Naturalist 166:E177–E192.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.2455/suppinfo