DR. DANIEL F. DOAK (Orcid ID : 0000-0001-8980-3392)
DR. ALLISON MARIE LOUTHAN (Orcid ID : 0000-0002-2735-6539)
DR. NATHALIE I. CHARDON (Orcid ID : 0000-0001-9120-4778)
DR. ROBERT K. SHRIVER (Orcid ID : 0000-0002-4590-4834)

Article type : Article

Journal: Ecological Monographs
Manuscript type: Article

Running Head: Comparing demographic modeling methods

Title: A critical comparison of integral projection and matrix projection models for demographic analysis

List of Authors:
Daniel F Doak*, Environmental Studies Program, University of Colorado, Boulder.
Ellen Waddle, Environmental Studies Program and Department of Ecology and Evolutionary Biology, University of Colorado, Boulder
Ryan E. Langendorf, Cooperative Institute for Research in Environmental Sciences and Environmental Studies Program, University of Colorado, Boulder
Allison M. Louthan, Division of Biology, Kansas State University, Manhattan KS and Biology Department, Duke University, Durham NC

Nathalie Isabelle Chardon, WSL Institute for Snow and Avalanche Research SLF, Flüelastrasse 11, Davos Dorf, Switzerland

Reilly R. Dibner, Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming;

Douglas A. Keinath, Wyoming Ecological Services Field Office, United States Fish and Wildlife Service, 5353 Yellowstone Road, Suite 308A, Cheyenne, Wyoming 82009; Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming

Elizabeth Lombardi, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY

Christopher Steenbock, Department of Ecology and Evolutionary Biology, University of Colorado, Boulder

Robert K. Shriver, Department of Natural Resources and Environmental Science, University of Nevada, Reno

Cristina Linares, Department of Evolutionary Biology, Ecology and Environmental Sciences, Institut de Recerca de la Biodiversitat (IRBio), University of Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

Maria Begoña Garcia, Pyrenean Institute of Ecology (CSIC). Avda. Montañana 1005 50059 Zaragoza (Spain) 50059 Zaragoza (Spain) Department of Evolutionary Biology, Ecology

W. Chris Funk, Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO

Sarah W. Fitzpatrick, W.K. Kellogg Biological Station, Michigan State University, Hickory Corners, MI

William F. Morris, Department of Biology, Duke University, Durham, NC

Megan L. Peterson, Plant Biology Department, University of Georgia, Athens, GA

*Corresponding author: daniel.doak@colorado.edu

ABSTRACT

Structured demographic models are among the most common and useful tools in population biology. However, the introduction of integral projection models (IPMs) has caused a profound shift in the way many demographic models are conceptualized. Some researchers have argued that IPMs, by explicitly representing demographic processes as continuous functions of state variables such as size, are more statistically efficient, biologically realistic, and accurate than classic matrix projection models, calling into question the usefulness of the many studies based on matrix models. Here, we evaluate how IPMs and matrix models differ, as well as the extent to which these differences matter for estimation of key model outputs, including population growth rates, sensitivity patterns, and life spans. First, we detail the steps in constructing and using each type of model. Second, we present a review of published demographic models, concentrating on size-based studies, that shows significant overlap in the way IPMs and matrix models are constructed and analyzed. Third, to assess the impact of various modeling decisions on demographic predictions, we ran a series of simulations based on size-based demographic data sets for five biologically diverse species. We found little evidence that discrete vital rate estimation is less accurate than continuous functions across a wide range of sample sizes or size classes (equivalently bin numbers or mesh points). Most model outputs quickly converged with modest class numbers ($\geq$ 10), regardless of most other modeling decisions. Another surprising result was that the most commonly used method to discretize growth rates for IPM analyses can introduce substantial error into model outputs. Finally, we show that empirical sample sizes generally matter more than modeling approach for the accuracy of demographic outputs. Based on these results, we provide specific recommendations to those constructing and evaluating structured population models. Both our literature review and simulations question the treatment of IPMs as a clearly distinct modeling approach or one that is inherently more accurate than classic matrix models. Importantly, this suggests that matrix models, representing the vast majority of past demographic analyses available for comparative and conservation work, continue to be useful and important sources of demographic information.

Key Words: Integral projection model, IPM, matrix projection model, demography, structured population, lambda, elasticity, life span

INTRODUCTION

Demographic models have yielded profound insights in many areas of ecology and evolution, including life history theory, population dynamics, resource management, and conservation biology. Studies using demographic approaches include some of the most influential papers in ecology (e.g., Cole 1954, Gillespie 1977, Shaffer 1981, Lande 1982, Pulliam 1988). Primary reasons for this influence are the ability of demographic models to link short term individual performance to both lifetime fitness and population growth. In addition, these models facilitate broad comparisons that can highlight trade-offs and limitations that structure diverse life history patterns (Stearns 1992). Standardized metrics from demographic models have spurred the development of general classification frameworks, such as the fast-slow continuum or the survival-growth-fecundity triangle (Silvertown et al. 1993, Franco and Silvertown 1996, Sæther and Bakke 2000, Gamelon et al. 2014, Salguero-Gómez et al. 2016b). Demographic modeling has also transformed approaches to conservation by allowing more quantitative assessments of population risk and potential management strategies (Schemske et al. 1994, Carroll et al. 1996, Biek et al. 2002, Morris and Doak 2002, Jongejans et al. 2008, Doak et al. 2015). Some of the most influential management plans for threatened, invasive, or economically important species have used demographic models to target specific life-history stages (Crouse et al. 1987, McEvoy and Coombs 1999) or to quantify the risk of extinction (Shaffer 1983, Lande 1988). Thus, demographic models are a cornerstone of both population biology and conservation management (Shea 1998, Caswell 2001, Morris and Doak 2002).

While demographic analyses need not involve distinctions between different types of individuals, the majority of such studies, even of annual species, fall within the realm of "structured" population models. These models are structured in the sense that individuals are classified by one or more "state variables," traits that are used to distinguish between individuals that are believed to have different demographic fates. State variables typically include age or size, but can also include many other predictors of fate, including life history stage, sex, microhabitat, or even symbiotic relationships or pathogen load (e.g., Palmer et al. 2010, Wilber et al. 2017).

While the underlying approach of demographic modeling has remained largely the same since the work of Leslie (1945) and Lefkovitch (1965), over the last two decades there has been a slow revolution in how many demographic models are conceptualized, symbolically presented, fit, and, to a lesser extent, interpreted. This is particularly true when species are described by one or more continuously varying state variables, many of which are descriptors of individual size. In these cases, integral projection models (IPMs), which describe populations according to a continuous state variable (Easterling et al. 2000), have begun to replace classic matrix models that begin by explicitly dividing populations into discrete categories corresponding to ranges of state variable values (Caswell 2001). The acknowledgement and understanding that discretizing continuous measures of size or other descriptors of state is a simplifying assumption of convenience goes back to the first uses of size-based demography in ecology (Vandermeer 1978, Moloney 1986), but the IPM literature has revived discussion of this simplification and suggested it is of paramount importance. The development and widespread adoption of IPMs has been motivated in part by arguments that a continuous approach is more biologically realistic and statistically efficient, particularly when applied to limited data sets (Easterling et al. 2000, Ellner and Rees 2006, Zuidema et al. 2010, Ozgul et al. 2012). Correspondingly, matrix models have been increasingly criticized as artificial, statistically inefficient, and prone to bias (Ramula et al. 2009, Salguero-Gómez and Plotkin 2010, Picard and Liang 2014).

In the last few years, several reviews have emphasized the superiority of IPMs (Merow et al. 2014, Rees et al. 2014); these claims have not, however, been critically or thoroughly evaluated. Previous tests of the relative accuracy of IPMs have been limited to comparisons that have ignored the multiple aspects of model estimation and development that can be used to formulate demographic models, and have also compared IPMs only with matrix models built with extremely few classes (Ramula et al. 2009). Further, most claims for the superior representation of the biology of species, due to the avoidance of artificial stage classes, ignore that in their actual implementation virtually all IPMs are analyzed as moderate- to high-dimension matrix models (Ellner and Rees 2006, Merow et al. 2014a). This means that their structure differs more quantitatively than qualitatively from traditional matrix models. Finally, there are potentially important biological simplifications inherent in the IPM approach that have not been carefully examined in the ecological literature, in particular the limitations imposed by representing vital rates as fairly simple continuous functions of the state variable.

An important consequence of the discussion surrounding the accuracy of IPMs and classic matrix models is that past studies using older methods could be perceived as providing little to no useful demographic information. Traditionally fit matrix models represent the vast majority of demographic data available for comparative studies (e.g., those in the COMPADRE and COMADRE databases; Salguero-Gómez et al. 2015, Salguero-Gómez et al. 2016a) or with which to assess the viability and management of species of conservation concern. Indeed, even as the accuracy of matrix models has been questioned, many synthetic reviews that reanalyze matrix models have recently appeared in the literature (Katz 2016, Csergo et al. 2017, Yokomizo et al. 2017). Thus, it is important to more carefully evaluate whether and when matrix models accurately capture population dynamics, and under what circumstances IPMs may do so with less bias and more precision. Finally, some of us (Doak, Morris, Garcia *pers. obs*.) have seen an increasing tendency of reviewers and editors to dichotomize these two approaches, considering matrix models as out of fashion and failing to recognize the considerable grey zone between the two model types as well as some of the subtler advantages and disadvantages of each.

Our goals in this paper are to i) explain the ways that matrix models and IPMs do and do not differ in their implementation, ii) articulate the potential pitfalls and advantages of each approach, and iii) use simulations based on real data sets to critically assess which model fitting decisions do and do not matter for common demographic outputs. To accomplish these ends, we start with an outline of the main steps in fitting matrix models and IPMs, highlighting the similarities and differences between these two approaches. Second, we review the demographic literature to document how population biologists fit these different models in practice, with the goal of evaluating how distinct they really are. Third, we present results from an extensive set of simulations based on five real demographic data sets. We use these simulations to compare the relative accuracy of matrix models and IPMs across a range of sample sizes, model-fitting strategies, and matrix dimensions that reflect the diverse approaches used by biologists. We conclude with a discussion of the merits and potential limitations of different demographic modeling strategies and recommendations for future demographic work.

TWO APPROACHES TO FITTING DEMOGRAPHIC MODELS

Both matrix models and IPMs seek to represent demographic heterogeneity within a population due to variation in individual state variables, such as age or size, that influence performance. Matrix models have traditionally been approached with the assumption that individuals can reasonably be divided into classes (also called categories, stages, or bins; we use "classes" in the subsequent text). These classes are based on subdivisions of the state variable, even when there is clear understanding that the underlying state variable is continuous (Hartshorn 1975, Vandermeer 1978, Moloney 1986). In contrast, IPMs explicitly seek to treat state variables as continuous.

In this section, we describe the four steps in formulating either kind of model: i) characterizing individual states, ii) estimating fates, iii) assembling these estimates of fates into a full demographic model, and iv) generating outputs from these models to assess individual fitness or population behavior. In Figure 1, we diagram these basic model-fitting procedures for "classic" matrix models and IPMs, illustrating the steps they share and those that differ between these approaches. We also use this diagram and discussion to highlight differences in the terminology used in both modeling strategies, as well as the similarities that can be disguised by these notational differences. Following sections on the four steps in demographic analysis, we discuss in more detail some of the features that most separate IPM and matrix models, and also some of the less obvious issues with using either approach.

Characterizing individual states

Both methods require the same basic demographic data: individual-level survival, growth, and reproduction rates, recruitment data, and one or more state variables that capture heterogeneity in these rates; these are called demographic or vital rates (Caswell 2001, Morris and Doak 2002, Franco and Silvertown 2004; note that some authors use vital rates to refer only to survival and reproduction). State variables may be chosen *a priori* based on feasibility or natural history, or selected from multiple variables by comparing regressions of vital rates on alternative state variables to find the ones with the highest predictive power (Morris and Doak 2002). In a matrix model, a state variable is either already discrete (e.g., age classes for a sharply seasonally-breeding species or the discrete life-history stages of many arthropods) or is divided into discrete categories of a continuous state variable (e.g., size classes). In the latter case, there are several algorithms for choosing the number and boundaries of classes (Vandermeer 1978, Moloney 1986), although in practice the structure of most matrix models has been decided based on natural history and data exploration (e.g., looking for sharp changes in vital rates; Caswell 2001, Ramula et al. 2020). In an IPM, the primary state variable is regarded as continuous, although additional discrete state variables such as age, sex, seedling state, dormancy state, breeding status, or others can also be included (Ellner and Rees 2006, Rees et al. 2006, Williams 2009, Jacquemyn et al. 2010). We note that age is commonly treated both as continuous and discrete in demographic models. This depends on data availability and whether reproduction occurs during well-defined time periods, generating discrete cohorts (i.e., birth-pulse), or offspring are produced more continuously throughout the year. However, state variables that reflect some aspect of size are most often used in IPMs (Figure 5) and can give rise to the largest differences in IPM and matrix model treatments, so we concentrate on these throughout the rest of the paper.

Characterizing individual fates

Both matrix models and IPMs use state variables to capture variation in vital rates – the fates of individuals, based on their state. While there are many ways to make such estimates, all fall into two general approaches. The first is used in the construction of many matrix models, where a separate estimate of each vital rate is required for each discrete class. These estimates come from separately analyzing the subset of data falling within that class to calculate, for example, mean survival or mean reproductive output. In other words, the vital rate for a given class is estimated independently of the rates for other classes. This includes approaches that take the observed transition frequencies for a given class as well as methods that fit statistical models that treat class as a categorical variable (e.g., some mark-recapture analyses). The central problem facing parameterization under this strategy is that more, narrower classes reduce the amount of data available for estimating each vital rate, whereas fewer, broader classes pool together individuals that may have very different fates. This trade-off has long been recognized (Vandermeer 1978, Moloney 1986, Ramula and Lehtilä 2005), and can mean that multiple iterations are needed to find a model structure that balances sampling and estimation error.

The second approach to estimating fates is used in the construction of some matrix models based on continuous state variables, as well as all IPMs. In this approach, demographers use continuous regression models of vital rates, fit to the entire data set, to estimate stage-dependent vital rate functions. Researchers generally use established functional forms for each vital rate: generalized linear models with binomial errors are often used for vital rates that inherently represent probabilities (i.e. survival, dormancy, or flowering) while those with Poisson or negative binomial errors have frequently been used for offspring numbers; general linear models have typically been used for growth rates . In some cases, splines or generalized additive models have been used to represent more complex relationships between vital rates and state variables (Dahlgren et al. 2011). Functions may be chosen *a priori,* or model selection methods, such as AIC, may be used to select from among several candidate models (e.g., linear vs. quadratic functions of state). One key difference between this strategy and the direct use of discretized data for vital rate estimation is the elegant way that size changes are treated

(Easterling et al. 2000). First, a model is chosen to characterize the mean size at the end of a time interval, given a starting size. Next, the squared residuals from this relationship are predicted in a second model and then used to predict the variance in ending sizes (although both the mean and variance can be fit simultaneously as well; Ellner and Rees 2006, Ellner et al. 2016).

Assembling a projection model

When building matrix models, the matrix elements are constructed from the best vital rate values for each class, either using the discrete vital rate estimates or an estimate for each size class taken from a continuous vital rate function (Batista et al. 1998, Morris and Doak 2002, Gross et al. 2005). This is straightforward if the vital rate estimates are made discretely for each class. If continuous functions have been estimated, different rules can be used to estimate the average value of a vital rate that is applied to a size class. Most commonly, the vital rate estimate corresponding to the midpoint size in the class is used, but other approaches, such as the vital rate of the mean or median size of individuals falling within a class, can also be employed (Morris and Doak 2002). In either case, the growth, survival, and reproductive rates estimated for each class are combined to form the elements of the matrix, a_{ij}, which represent the average number of individuals in class i at time t+1 that result from an individual of class j at time t.

In IPMs, most vital rates are estimated by fitting continuous functions of one or more state variables. When building an IPM, these fitted functions are then combined into density kernels. These are usually a survival/growth kernel that describes the distribution of an individual's state in the next time step, given survival and growth, and a reproduction kernel that describes the number and state distribution of an individual's offspring. These kernels are then combined into an overall kernel that projects the number and distribution of individuals' states across a time step. In this kernel, k_{ij} is identical in interpretation to the matrix element a_{ij}, except that the i and j states are assumed to apply to size classes for the matrix model and to point values of the state variable for IPMs. Proponents of IPMs emphasize that this regression-based approach avoids artificial binning together of individuals with differing states and, by including all individuals in the model fitting step, allows more efficient use of scarce data (Easterling et al. 2000, Ellner and Rees 2006, Ramula et al. 2009, Zuidema et al. 2010, Merow et al. 2014b).

Beyond these common ways of building either type of model, several other complexities and complications can arise. Most commonly, additional state variables (e.g., sex, age class, widowing status; Miller and Inouye 2011, Bakker et al. 2018) or other covariates (e.g., climate, soil chemistry; Dahlgren and Ehrlén 2009, Doak and Morris 2010, Hunter et al. 2010, Diez et al. 2014, Merow et al. 2014b) may have important effects on individual fates and can be included in either discrete or continuous approaches to vital rate estimation (e.g., through additional or combined classes, or by inclusion in continuous vital rate functions). In addition, both matrix models and IPMs can be either deterministic or stochastic, including the influence of demographic and environmental stochasticity on vital rates to estimate effects on fitness, growth rates, or extinction risk. There are also increasingly sophisticated methods to incorporate model and parameter uncertainty into the predictions of these models, which is an especially important topic when models are being used to address applied questions (Bakker et al. 2009, Elderd and Miller 2016). Finally, it is worth noting that there are many other subtle and not-so-subtle decisions that must be made when formulating either a matrix model or an IPM (detailed in Caswell 2001, Morris and Doak 2002, and Ellner et al. 2016), and numerous mistakes are commonly made in model construction. For example, a recent review by Kendall et al. (2019) found that a substantial fraction of matrix models constructed for animals contained at least one common error in model structure. Among the most common of these mistakes are failing to include survival in reproductive rates, introducing incorrect delays into the life history, and incorrectly calculating transition rates from stages with known duration (Kendall et al 2019).

Analysis of Demographic Models

Once constructed, matrix models are used to compute multiple biologically important outputs. Most commonly, these include one of several measures of population growth rate, including asymptotic or transient measures of deterministic or stochastic population growth (Caswell 2001). Additional outputs include the stable stage distribution, damping ratio, life span measures, and the sensitivity and elasticity of population growth or of other outputs (e.g., stable stage distributions; Caswell 2001, Morris and Doak 2002, Haridas and Tuljapurkar 2005) to either matrix elements or vital rates. If models are built with continuous vital rate functions, sensitivities can also be estimated for responses to changes in parameters of vital rate functions (e.g., the intercept or slope of reproduction as a function of size) rather than to discrete class-specific values (Griffith 2017). Caswell (2001) provides a thorough review of the many outputs of matrix models, and multiple computing packages facilitate these analyses (e.g., popbio in R; Stubben and Milligan 2007).

How does one get comparable predictions from an IPM, which is not a matrix, but a density kernel? IPMs are actually analyzed in the same way as matrix models, using discretized matrices, although IPM nomenclature often obscures this fact. In practice, numerical integration methods are used to approximate an IPM kernel as a transition matrix, most often based on discrete "mesh points," which are starting and ending values of the state variable. This analysis method divides the state variable into many classes, centered on the mesh points, within a biologically plausible range and then uses the values of each of the different vital rate functions at each mesh point to estimate the transition rate from each class to each other class. The result is a moderately-sized to large matrix with many narrow, discrete classes and transition rates estimated from the vital rate functions underlying the IPM kernel. It would be possible to analyze IPMs without discretization, but it would be a far more formidable analytical challenge for arbitrarily defined kernels (Ellner et al. 2016), while the methods of linear algebra make the analysis of the approximating matrix straightforward.

Both modeling approaches result in large to very large numbers of certain outputs, such as sensitivity and elasticity values to size specific vital rates or matrix elements. To deal with these sometimes daunting numbers of values and to provide more succinct and biologically informative results, for both types of models practitioners frequently condense results into mean or summed values for fewer categories (e.g., Zuidema et al. 2010, Silvertown et al 1993).

Comparing the two approaches

The IPM literature has emphasized two shortcomings of matrix models: 1) the statistical inefficiency of separately estimating vital rates for each class, and 2) the use of a small number of classes to represent inherently continuous state variation, a situation that can lead to mischaracterization of the true values of individual fates. IPMs solve these problems by using all individuals to estimate continuous vital rate functions and then by using many classes of small width in the final analysis phase. However, there are reasons to question whether matrix models and IPMs are truly as different as they are usually portrayed. First, as noted above, demographers have used continuous vital rate functions to characterize patterns in vital rates and to parameterize matrix models, even well before IPMs were developed (e.g., Siler 1977, Eberhardt 1985, Barlow and Boveng 1991,Batista et al. 1998, Bernal 1998, Zuidema 2000, Morris and Doak 2002, Matsuda and Nichimori 2003, Gross et al. 2005, Rogers-Bennett and Rogers 2006, Chien et al. 2008). Second, in practice IPMs are analyzed by discretizing the underlying kernel to parameterize a matrix model, although that matrix is usually moderately to very large (typically many dozens to hundreds of classes). Thus, rather than describing matrix models and IPMs as completely distinct methods, it is more accurate to view structured population models as varying along at least two axes: the method of parameter estimation (categorical vs. continuous functions, blue boxes in Figure 1) and the dimensionality of the resulting matrix (number of classes or mesh points, green boxes in Figure 1). Whether these two frameworks are distinct or not, the problems that have been identified with discrete parameterization and with modeling continuous state variables with few classes can potentially have serious effects on model predictions. In the rest of this section, we briefly review important considerations arising from these two aspects of model-fitting, as well as several other potential issues.

Class number. IPM practitioners seek to reduce the effects of discretization by using many narrow classes (Merow et al. 2014a). However, many classes make for larger matrices and reduce computational efficiency, especially when there are multiple state variables (Ellner et al. 2016). There is also little information about the number of classes necessary to approximate a continuous demographic process for real life histories. Several iterations may be required to determine the number and range of classes needed to yield stable estimates (Ellner and Rees 2006, Zuidema et al. 2010) and avoid eviction (the removal of individuals from the range of model sizes due to esimates of growth or shrinkage outside this range; Williams et al. 2012). The fact that most matrix models have far fewer classes than the large matrices used by IPMs has been viewed as a key advantage of IPMs (Ramula et al. 2009, Zuidema et al. 2010, Merow et al. 2014a). However, in the few studies of which we are aware that test for class number effects in IPMs, increasing classes beyond 10 to 20 has little effect on model results (e.g., Jacquemyn et al. 2010, Shriver et al. 2012, Dibner et al. 2019).

One factor that is likely to influence the number of classes needed for accurate or stable predictions is the way that continuous vital rate functions are discretized. The most common approach in IPMs is to use the “midpoint rule” to evaluate the vital rate functions across mesh points, each representing the midpoint of a class of the state variable, to obtain point estimates of the survival and fecundity rates that contribute to the k_{ij} values in the discretized kernel. There are two important variations on this method that directly relate to the number of mesh points (or, analogously, classes) necessary for a reasonable approximation. First, it has been suggested that it may be more accurate to characterize the vital rates of a class by using either the median or mean state value (Morris and Doak 2002) of the individuals in a dataset falling within a class, rather than the midpoint. A similar approach is to estimate an empirical density function for individual states to estimate mean or median values, which can provide estimates even for classes in which no individuals were censused (Gross et al. 2005).

A second issue is the way that the transition probabilities between state values (e.g., growth and shrinkage probabilities for a size-based model) are discretized (Figure 2). The most commonly-used approach in the IPM literature (Ellner and Rees 2006, Metcalf et al. 2013, Merow et al. 2014a, Elderd and Miller 2016) approximates the probability density function (PDF) describing state at the next time step, conditional on starting state, by evaluating the probability density at each mesh point and then multiplying this value by the class width (Figure 2B). A more accurate method, but one that is not featured in most descriptions of IPMs or in the software to run these models (e.g., IPMpack; Metcalf et al. 2013) is to use the cumulative density function (CDF) to integrate the probability density across the entire class (Figure 2C). Although these two methods will converge with infinitely many classes (Fig. 2D), the first may require many more classes to produce stable estimates, particularly if the variance in size is small relative to the width of the classes for at least some starting sizes (Fig. 2E; Ellner et al. 2016). Although the second method has been used, including by the authors (e.g., Louthan et al. 2018, Montero-Serra et al. 2018), in both matrix and IPM models, the method of discretizing individual changes in state (e.g., growth) is virtually never reported in the methods of published studies (pers. obs.). We thus have no quantitative estimate of the relative frequency of these two approaches, despite their potential to influence the accuracy of model predictions. We also note that there is a third option to discretize size transition data, the “bin-to-bin” method (sect. 6.8; Ellner et al. 2016), which uses the integral over both the starting and ending sizes included in a transition to estimate total transition probability.

*__Vital rate estimation__***.** We next consider other aspects of using continuous vital rate (CVR) functions vs. discrete vital rate (DVR) estimates that may be less obvious, but are important in generating accurate models. The statistical advantages of fitting continuous functions are clear: using all individuals to fit a single function is more efficient than separately estimating vital rates based on a subset of individuals within each of many classes. In particular, it has been argued that this approach is more accurate than using discretely estimated rates in matrix models, given small datasets (Ramula et al. 2009). Discrete estimation also means that outliers or other quirks in the finite data used may have undue influence on the model structure and predictions (e.g., estimating zero or perfect survival for some classes).

On the other hand, there is also a potential cost of continuous vital rate function estimation that has received less attention. The functions used to explain vital rate variation are often quite simple, usually linear or perhaps quadratic functions of a single state variable (Merow et al. 2014), and thus can easily oversimplify or misrepresent how vital rates vary as a function of the state variable. In contrast, matrix models that separately estimate vital rates within each category make no such distributional assumptions (Shimatani et al. 2007). For example, a matrix model can estimate sharp discontinuities in survival probabilities between size classes or survival rates that asymptote well below 1, whereas IPMs usually model survival as a smooth logistic function of size with an asymptote of 1 (see Yau et al. 2014). Simple transformations of size variables, such as logging, may solve some, but not all, of these issues. For these reasons, some have advocated using non-parametric methods (Ellner et al. 2016, section 10.1.5) or fitting more complex functions, such as splines. However, these approaches can also be influenced by outliers and/or low sample sizes at extreme state variable values (Shimatani et al. 2007, Dahlgren et al. 2011, Rees et al. 2014). A related issue when using either approach is how best to account for estimation uncertainty and thus isolate process variance in vital rates.

Another issue with vital rate estimation is the ubiquitous assumption in IPMs of normally distributed growth rates on the scale of the state variable (Peterson et al. 2019). This assumption means that growth is modeled as symmetric around an average size transition, whereas for many species the distribution of growth is skewed. For example, high shrinkage may be more likely than high growth due to die-back, breakage, or starvation (reviewed in Peterson et al. 2019); the opposite pattern may occur in woody species measured using diameter at breast height (Needham et al. 2018). More generally, the use of growth models with infinite tails, like normal distributions, will predict some chance of growth and shrinkage to sizes well outside the range of reality, resulting in the problem of eviction (Williams et al. 2012) as well as unrealistic changes in state even within the bounds of otherwise realistic sizes.

There are multiple statistical methods to account for any of the complexities just discussed, but very few empirical demographic studies employ these. In addition, it is important to note that the goal of model development is *not* to represent every nuance of reality, but instead to get the important aspects right enough to yield useful representations of the patterns and dynamics of interest. But, as this perspective emphasizes, neither matrix model nor IPM approaches are *a priori* more compelling than the other. Both make some simplifications and smooth over some patterns in the data, though they do so in different ways. It is not clear whether and under what circumstances continuous functions will produce more accurate vital rate estimates compared to separately estimating vital rates within discrete classes, nor is it clear when and why the use of many narrow classes will fundamentally change model predictions. With this in mind, we next turn to how the two modeling approaches have actually been used in recent demographic studies.

HOW ARE IPMS AND MATRIX MODELS USED IN PRACTICE?

As we argue above, IPMs and matrix models are not sharply distinct. Here we document the range of methods used to fit these models in the literature, including different parameter estimation approaches and matrix dimensions. We conducted a literature search on October 23, 2018 of studies included on Web of Science using the search terms "demograph*" and "matrix" and either "ecology" or "conservation" for the period 2002-2018. We believe that these search terms capture the vast majority of matrix models used for ecological or life history analyses. We also included Web of Science results that had cited any of the papers originally developing the IPM approach, including Easterling et al. (2000), Ellner and Rees (2006), and Rees and Ellner (2009); because terminology, and hence keywords, are less uniform for IPMs, we felt that using citations of these founding articles would capture studies that would otherwise be missed. The starting year for our review is somewhat arbitrary, but was chosen to include virtually the entire period during which IPMs have been conducted. We only included papers that fit new models to demographic data, excluding strictly theoretical papers or review papers that relied on previously published models. For each paper, we determined the state variable (age, size, stage, or other [including size x age models]), method of parameter estimation (categorical, continuous, or a combination), the type of model as it was identified by the authors (matrix vs. IPM, stochastic vs. deterministic), and the dimension of the resulting matrix (number of classes, bins, or mesh points used to construct the matrix or discretized IPM kernels). For papers with multiple species, we identified these criteria for each species separately. Ambiguous papers were reviewed by at least two people. We identified 794 publications and 1271 demographic models across a range of taxonomic groups (Table 1) that fit all of our criteria. Most of the demographic studies in our database examined plants and other autotrophs (N=698), followed by vertebrates (N=486) and invertebrates (N=87). The full results of this literature review are available in Data S1: Literature Review.

Matrix models represent the majority (~79%) of demographic models published between 2002–2018 (Table 1). Over this period, 57% of all studies were deterministic matrix models, followed by stochastic matrix models at 22%. 21% of demographic models were identified by the authors as IPMs (16% deterministic IPMs, 5% stochastic IPMs). Although IPMs are a smaller fraction of published demographic models, this proportion has increased over time (Figure 3A; logistic regression of proportion of models: year coefficient = 0.31, Z = 12.81, P < 0.001). This increase can be attributed to the publication of several reviews of the method (Ellner and Rees 2006, Rees and Ellner 2009, Merow et al. 2014a, Rees et al. 2014) as well as the development of IPMpack, an R package for constructing IPMs (Metcalf et al. 2013, R Core Development Team 2015).

Almost all IPMs are built for size-based models. 67% are only structured by size and an additional 29% use size in conjunction with one or more other state variables (e.g., birth date, age, growth rate, dormancy, developmental stage, etc.). 26% of IPMs use age as a state variable, while only 2% use something other than age or size as the primary state variable (e.g., infection load; Wilber et al. 2017). In contrast to IPMs, matrix models are commonly used for stage and age-based analyses as well as size-based models; 20% use a size-based state variable, 28% use age, and 52% use a measure of stage (Appendix S1: Figure S1A,B). In addition, the great majority of published IPMs have been for perennial plant studies, while a wider range of taxa and life histories have been the subjects of matrix models (Appendix S1: Figure S1C,D). Given these differences, in the rest of our review we concentrate on comparisons between matrix models and IPMs that are based on size. The time trends of just these studies is similar to those of all demographic models (Figure 3).

In general, size-based IPMs used higher dimension matrices in their final analyses than matrix models (Figure 4; linear model of class number: t = 10.65, P < 0.001). However, IPM papers reported using an astonishingly wide range of classes, from 39 to 2400, to discretize their projection kernel (mean = 242.5, median = 200, SD = 308.3, N = 83) whereas size-based matrix models varied from 2–67 classes (mean = 7.3, median = 5.5, SD = 6.7, N = 194). However, only 43% of IPM studies reported the number of classes or mesh points used for the discretization of the kernel. This may reflect the philosophical view that IPMs should be conceptualized as continuous, despite their ultimate discretization, or may simply be viewed as an unimportant detail by those publishing results of these models. Regardless, we could not determine the number of classes for over half of published IPMs, although we note that the R package IPMpack uses 50 classes as the default setting (Metcalf et al. 2013) and 10.3% of IPMs reviewed that did not report the number of classes used IPMpack. We were further unable to determine the method used to discretize the IPM kernel for over a third (36%) of published IPMs. Of those that reported the discretization method, 96% used the midpoint rule and only 4% used an alternative integration method (e.g., Simpson's Rule, Gauss-Legendre quadrature; Ureta et al. 2012, White et al. 2016, Molowny-Horas et al. 2017). Given that the lack of discretization is frequently discussed as an advantage of IPMs, but that models are in fact analyzed with discretization, we urge that discretization information should be included in the description of any IPM analysis, since this is a key analysis step for these models.

Population biologists followed a variety of workflow paths from data to final matrix analysis (Figure 5). Most demographic models used information on individuals' stages to estimate vital rates categorically and construct deterministic matrix models (Figure 5). However, 7% of all matrix models, and 25% that used size as their state variable, estimated at least one vital rate using a continuous function of state. Interestingly, the proportion of matrix models using continuous vital rate estimation appears to have peaked and then declined over time (Fig. 3B; logistic regression of proportion of models: year coefficient = 141.89, z = 3.69, P< 0.001, year^2 coefficient = -0.035, z = -3.69, P< 0.001). This could reflect an increased awareness of continuous approaches to vital rate estimation following examples in Morris and Doak (2002) and the initial development of IPM methods (Easterling et al. 2000, Ellner and Rees 2006), with a later decline as IPMs were increasingly adopted to model data sets suitable for continuous vital rate estimation. When comparing size-based matrix models, we found that models tended to use more classes when at least one vital rate was estimated using a continuous function (mean = 11.0, median = 7, range = 3-67, N = 48) relative to models with discrete vital rate estimation (mean = 6.1, median = 5, range = 2-27, N = 146; linear model of class number: t = 4.65, P < 0.001). However, we found no relationship between class number and minimum sample size for discrete size-based matrix models (r = 0.17, P = 0.16, N = 67; Appendix S1: Figure S2). Of all demographic models using continuous vital rate estimation, 22% were matrix models. Thus, any advantages of continuous vital rate estimation have not been limited to IPMs in the demographic literature.

ASSESSING THE CONSEQUENCES OF DIFFERENT MODEL-MAKING DECISIONS

Two key characteristics of structured population models, the number of classes into which the state variable is divided and the method of parameter estimation (continuous vital rate functions (CVRs) vs discrete vital rate estimates (DVRs)), are often assumed to covary between "pure" matrix vs. "pure" IPM approaches. But as we show in our literature review, they are not necessarily logically connected and many published models combine relatively small class numbers with CVR functions.

There are also three other decisions that require careful thought when turning data into a structured demographic model, but are rarely discussed. First is the exact way that class boundaries are delineated. For most IPMs, class boundaries are set at regular intervals, while for matrix models, there are often decisions made regarding sample size issues and where size breaks make the most biological sense. A second decision is how best to characterize the average vital rate value for a given class or, using IPM terminology, how to define the mesh points used to evaluate the CVRs to create a discretized matrix. Mesh points are most often chosen as the midpoint of a class, but alternative approaches could use the mean, median, or even distribution of state values observed in each class. The third consideration, if using continuous functions, is how the transition probabilities between states conditioned on survival (e.g., growth) are discretized. This is distinct from the discretization of other vital rates, such as survival or fecundity, because an individual will have a distribution of possible states at the next time step (vs. a point estimate of survival probability or offspring number), and it is this continuous probability density that must be discretized (Fig. 2A). One approach is to use the point estimate of the probability density evaluated at each mesh point, multiplied by the class width (Figure 2C). Alternatively, the probability density can be integrated across the entire range of states within each class, by taking the difference between cumulative distribution function (CDF) values at the upper- and lower-class boundaries (Figure 2A, B). Other approaches, including the Ellner et al. (2016) "bin-to-bin" method or direct parameterization of discretized growth probabilities (Shriver et al. 2019) can also be used, but none of these alternatives have been commonly employed to date. All of these decisions have the potential to interact with the number of classes and methods of estimating vital rates to shape model outputs.

To test how these different aspects of demographic modeling influence model predictions we used large demographic data sets from five diverse organisms – a long-lived subtidal Mediterranean gorgonian coral (*Paramuricea clavata*, Plexauridae), a long-lived rupicolous plant (*Borderea chouardii,* Dioscoreaceae), a moderately long-lived arctic/alpine geophytic plant (*Polygonum viviparum,* Polygonaceae), a short-lived epiphytic lichen (*Vulpicida pinastri,* Parmeliaceae), and a short-lived fish, the Trinidadian guppy (*Poecilia reticulata*, Poeciliidae). While these species do not span the entire range of life histories seen in plants, animals, and fungi, they do represent a broad array of key life history patterns. In all these data sets organism size is used to structure the populations, but the species differ in multiple aspects of their ecology and capture a range of size distributions (Figure 6).

In our analyses we varied five aspects of model construction. Most fundamentally, we varied the method used to estimate vital rates (blue boxes, Figure 1) and the class number of the resulting matrix (green boxes, Figure 1). Previous comparisons of matrix dimension and parameterization methods have only included matrix models with a few classes (4–6) and discrete parameter estimation, and compared them to IPMs using continuous functions discretized into large matrices (100 classes; Ramula et al. 2009). In contrast, we varied parameter estimation method independently from class number to ask how each affects model accuracy. In addition, we tested the effects of the three other modeling decisions just mentioned: (i) use of midpoint or estimated median individual sizes for CVR estimation of average vital rates per class; (ii) even or sample size-adjusted class boundaries; and (iii) the ways in which discretized growth probabilities were estimated from CVR models (Figure 2). While other issues also influence model structure and results (see Section 2, above), here we concentrate on this short list of issues that will influence virtually all models.

We tested the effects of these decisions on three common demographic outputs: estimates of individual fitness or population growth (lambda, λ), estimated individual longevity (age at which 1% of individuals starting in the smallest class are still alive), and damping ratio (the ratio of the magnitudes of the dominant and subdominant eigenvalues), a measure of the strength and duration of transient dynamics for populations not at a stable stage distribution (Caswell 2001). While multiple other measures of longevity and also of the strength and length of transient dynamics exist, these measures have been widely used in the ecological literature. We also present a more limited comparison of how sensitivities and elasticities vary as a result of different modeling approaches. Finally, we test how data quantity interact with these alternative modeling decisions, in particular asking if some modeling approaches are more robust when data are scarce.

Study species and data sets

We compiled data used in published studies for each of our study species, supplemented with some unpublished information needed to employ flexible and automated model fitting routines; all data were collected by the authors. While all of the original studies of our species included effects of temporal and/or spatial variation in demography, in our simulations we used all transition data at once to construct single deterministic models. This simplification allowed us to use large numbers of observations to construct single models, as well as to rarify our data sets over a wide range of sample sizes to check the effects of data quantity on the relative merits of different modeling strategies.

We briefly outline the most relevant information about each species' life history and details about the data and modeling protocols used here:

Polygonum viviparum, the alpine bistort (hereafter, bistort), is an arctic/alpine perennial plant for which demographic data were collected annually from 2001–2011 at four populations on Niwot Ridge in Colorado, USA (Doak and Morris 2010), for a total of 11,882 plant-transitions of data. Size, reproduction, and survival data were recorded for all plants (see methods in Doak and Morris, 2010). Size is measured as the square root of estimated leaf area in mm^2; on an untransformed scale, sizes in the main data set (not including recruit sizes) range from 4.39 to 3600, after truncating 5 large values that created a long sparse tail that created problems for some of our analyses (similar truncation was done for all data sets besides *Borderea*). Reproduction is exclusively via asexual bulbils that are produced on inflorescences, and our measure of reproduction is the size-dependent product of the probability of producing one or more inflorescences and the estimated number of bulbils produced if reproducing (derived from a contunuous measurement of the length of the inforescence bearing bulbils). We pooled data across all years and sites to yield one estimate of the number of new recruits (bulblings) seen a year later per bulbil produced (0.00676). Bulbling sizes were also pooled and used to characterize the size distribution of bulblings as normally distributed with a single mean (4.00) and SD (0.886) for CVR models, and were directly used to get frequencies of sizes for DVR models.

Paramuricea clavata, the Mediterranean red gorgonian (henceforth, gorgonian), is a slow-growing, long-lived arborescent octocoral that typically occurs from 15 to 60 m depth. Demographic data for individual colonies (the unit of demographic analysis) were collected annually at three Mediterranean sites for 2-4 years each from 1999–2004 (Linares et al. 2007, Linares and Doak 2010), for a total of 4,877 colony-transitions of data. Size is quantified as colony height; sizes in the main data set range from 0.2 to 74.9 cm. Size-dependent reproduction was estimated as the production of oocytes per colony, estimated from the relationship between gonad number and size derived from data in Coma et al. (1995; Table 5) and the estimated average oocyte number resulting from a gonad (2.77774×10^{-6}). We estimated a common first year survival of new recruits across all years and sites as 0.667, the mean of colony survival in the smallest size class from Linares et al. (2007). Surviving recruits were assumed to have a uniform size distribution ranging between 0.3 to 3.0 mm height.

Vulpicida pinastri (henceforth, *Vulpicida*), is a relatively short-lived epiphytic lichen that grows on several species of trees and shrubs. Data on individual thalli were collected annually from 2004-2009 in Kennicott Valley in Alaska, USA on individuals growing on *Alnus* stems in a mixed spruce-alder forest, for a total of 1621 individual transitions of data. Size and survival data were collected in each thallus, with the square root of thallus area in cm^2 used as the measure of size (see Shriver et al. 2012 for methods); on an untransformed scale sizes in the data set range from 0.15 to 47.61. Reproduction was estimated as proportional to the circumference of a thallus, which bears the majority of asexual propagules. The number of recruits per mm of circumference necessary to achieve a stable population was estimated as 0.047 in the original study of this species (Shriver et al. 2012), and we used this estimate as a fixed value in our models. New thalli sizes were estimated to have a uniform distribution ranging between 0.124 and 0.50, which reflect the range of smallest thallus sizes encountered in the field.

Borderea chouardii (henceforth *Borderea*) is a rare, extremely long-lived rupicolous plant that naturally inhabits a single population in Spain, where it grows in shaded crevices of north-facing limestone walls and overhangs. Data were collected on individual plants from 1995 to 2002 at two sites in the Spanish Pyrenees, for a total of 2,682 plant-transitions of data. Size is measured as the length of the largest leaf in mm; sizes in the main data set range from 2 to 10.8.

Reproduction is quantified as the number of seeds produced per female plant. While the species is dioecious, we use the mean seeds per plant of a given size, averaged across the sexes, as our measure of reproduction (see Garcia 2003 for more details of sampling and life history). Reproduction was quantified as the size-dependent product of the probability of producing one or more inflorescences and the estimated number of seeds produced if reproducing. For CVR models, observed seedling sizes were used to estimate a mean (4.66) and SD (0.46) and sizes were assumed to be normally distributed, while the set of observed sizes were directly used to get frequencies of sizes for DVR models.

Poecilia reticulata (henceforth, guppy) is a short-lived freshwater fish native to streams and rivers in Trinidad. We used capture-mark-recapture data from monthly sampling of a site on the Caigual River that spanned January 2009 to June 2011, for a total of 4880 unique individuals over the entire study. At each sampling interval, a comprehensive capture of all fish within the stream reach was attempted, with size (wet mass in g) and sex recorded for all individuals. Sizes in the data set range from 0.042 to 0.904. While recapture rates are extremely high (Fitzpatrick et al. 2016), they are not perfect. To produce a simplified data set for our analyses, we therefore considered a fish dead at the first census it was not captured, if it was not captured for at least one subsequent sampling period (thus, we did not include data from the final two sampling intervals). We also linearly interpolated size for fish that were not sampled in a month, but were sampled in the months bracketing the missing capture. We built a model for females only. Reproduction was quantified as the size-dependent product of the probability of producing any offspring times the number of daughters produced if there was reproduction. Offspring number was estimated from genetic data and is the estimated number of female offspring produced that survived until the second census following birth (newborns were not large enough to reliably catch until approximately 1-2 months of age; Fitzpatrick et al 2020). New offspring sizes were pooled and used to characterize an emprical distribution function for use in CVR models, and were directly used to get frequencies of sizes for DVR models. The survival rate of fish in each of their first two months was estimated as the mean survival estimated for newly observed fish, based on a logistic regression of monthly survival vs size fit to all fish. We do not include in our analyses data on the genetic origin or hybrid status of the fish (see Fitzpatrick et al. 2016 and Fitzpatrick et al. 2020 for more details of the study). As the dataset only allowed estimates of reproduction from months 5-14 of the study, we used demographic data only from these months, for a total of nine individual transitions of data and 2366 individual monthly transitions. In keeping with the data collection and also the limited life span of the species, and unlike the annual time steps used in the models for all other focal species, all data analysis and modeling for guppies was done using a monthly time step.

Methods

We fit demographic models to the data from each species using combinations of the following alternative approaches:

A. **Discrete vital rate estimation for each class vs. continuous functions for vital rate estimation.** When estimating continuous vital rate functions (CVRs), we fit separate size-dependent models for survival, mean growth, variance in growth, and the reproductive rates, described above for each species (see also Appendix S1: Table S1). For each vital rate, we fit 2–3 models with alternative size-dependent functions and used AICc to choose the best model. Specifically, we fit models with linear vs quadratic size effects for all vital rates, and for mean growth we also fit a power function, to potentially better capture different shapes of non-quadratic, but non-linear shifts in growth with size. These functions are all commonly used in analyses employing CVRs.

B. **Number of classes for model construction.** For discrete vital rate estimation, the number of classes directly influences the parameter estimation, while for continuous vital rate functions it only influences the final construction of the matrix for analysis. We made models with class numbers that ranged from 3 to 100 classes for most analyses, using 3, 4, 5, 6, 8, 10, 15, 25, 35...100 classes. With discrete vital rate estimation, at the upper end of this range we rapidly reached class numbers that resulted in low samples for at least one class; we did not make models if the smallest class-specific sample size was < 3. While this is a very lenient standard (we do not advocate making models based on such low sample sizes per class), we used a low threshold in order to make DVR models with the largest possible range of class numbers given the data we had. In our rarefaction tests, described below, we also directly tested the effects of having very low class-specific samples on DVR model outputs. Statistics on per class sample sizes for each species are given in Appendix S1: Figures S13-17.

C. **Proportional vs. even size class delineation.** We employed two approaches to defining boundaries of size classes, which characterize two extreme approaches seen in demographic studies. With even class divisions, all classes were the same width on the scale of the size metric (see above for definition of the size scale used for each species). For proportional class divisions, we used the *classes* function in R package binr (Sergei 2015) to create class divisions that had as nearly equal numbers of starting individuals as possible. This approach has the general effect of creating many narrow classes of smaller or mid-sized individuals and fewer wide classes for the larger and in some cases also small individuals, depending on the size distribution of the data (see Figure 6). While in many matrix models class boundaries are, and should be, made with more attention to biological breakpoints, to automate the process of choosing class boundaries we used only these two approaches.

D. **Discretizing CVRs.** For most vital rates, a single point estimate is required for each size class (e.g., survival probability, number of offspring/parent). We used one of two approaches to estimate the representative vital rate value for each size class when using the CVR approach. First, and most simply, we used the midpoint size within a class (the mean of the two bounding values for the class). This method is by far the most common one used when making large matrices to numerically integrate IPM models, and is also used in many matrix models employing CVRs. However, two of us have argued that it is more representative to use a size that reflects the average individual within a class, not the midpoint of the class boundaries (Morris and Doak 2002). This estimation can be accomplished in several ways. Most simply, an estimate can come by taking a simple median or mean starting value of all individuals within a class, or, when data are scarce in some size ranges, by fitting an empirical density function to all individuals in the population and then using this function to create a weighted median size for each class—we used this latter approach in our simulations. We refer to these two approaches as midpoint- or median-based CVRs, respectively.

E. **Discretizing continuous growth distributions.** Lastly, we compared two approaches to discretizing continuous distributions that summarize changes in size when using the CVR approach (Figure 2), using both simulated and real data sets. First, we took the common approach employed in the IPM literature of using the point estimates of the probability density for the midpoints of all of the size classes, multiplied by the class width, to approximate the probability density function (PDF) of size at the next time step conditioned on current size. We call this the “mesh point method.” Second, we used the cumulative density function (CDF) for growth to get the probability of reaching each size class at the next time step conditioned on starting size. We call this the “CDF difference method.” While other approaches have been proposed, in particular the Ellner et al. (2016) “bin-to-bin” approach, we only tested these two most commonly used approaches here. For either, there is a concern that some substantial fraction of the total probability of growth will fall outside the upper and lower limits for size defined in the model (the eviction problem; Williams et al. 2012). There are multiple ways to correct the estimated growth probabilities so that, for any starting size (or size class), they sum to one (Williams et al. 2012). We do so by renormalizing the growth probabilities for a given starting size by the difference of the CDFs for the minimum and maximum sizes used in the model (as in Williams et al. 2012). While this correctly standardizes total growth rates for the CDF difference method, it is a more error-prone exercise for the mesh point method, as we discuss in Results.

Appendix S1: Table S1 lists the vital rates fit for each species, including non-size-dependent vital rates used in all models. Appendix S1: Figures S3-7 show the best-supported vital rate functions for each species. As these figures show, the species span a range of patterns in growth, survival and reproduction. In particular, *Borderea* (Appendix S1: Fig. S4) and guppies (Appendix S1: Fig. S7) show a pattern of declining mean and variance in growth at larger sizes, while the other species show declining mean but increasing or relatively stable variance as size increases.

For each model constructed from each data set, we estimated λ, damping ratio, and longevity. We also explored the effects of class number and discrete vs. continuous vital rate estimation on elasticity values, contrasting DVR models with 20 evenly-spaced size classes with CVR models built with 80 classes and using median-based and CDF difference methods. 20

class models are at the upper range possible to use for simple even class definitions for all our data sets, while use of 80 classes is well within the range used by most IPM models (Figure 4).

Finally, we tested the effect of sample size on model results by rarifying each data set in two different ways. First, we randomly sampled each dataset, with replacement, 200 times for the full sample size, retaining ½, ¼, 1/8, down to 1/32th of the data, depending on species. We then replicated the model-fitting process for 20 class DVR and CVR models, employing median-based and CDF difference methods for discretization. For these models, we use a slight variant on even class boundaries. To perform analyses on rarified data sets using even size classes for DVR estimation, we had to use a stratified bootstrap approach, so that we retained some individuals across sizes. This stratification regime also reflects the empirical sampling decisions that many demographers make, with efforts to include individuals at the top and bottom of the size distribution in the sample followed for data collection. To stratify the sampling, and also to define class boundaries for the DVR, we used the top and bottom 5% of all individuals by starting size to define the smallest and largest classes. We then divided the remaining individuals into 18 even size class divisions. Bootstrapped samples were generated by resampling with replacement separately for each size class. Resampled data sets ranged from a sample equal to the original (see Appendix S1: Figures S13-17 for per class sample sizes statistics), down to between 1/8th and 1/32th of the original sample size, reflecting quite small minimum sample sizes of individuals in a given class: bistorts, 6; gorgonians, 5; *Borderea* and guppies, 3; *Vulpicida*, 2. As supplementary tests, we also: 1) fit 80 class CVR models to each data set to see if higher class numbers changed the results, and 2) ran similar rarefactions, but using non-stratified bootstraps and fitting proportional class boundary models using 20 classes for DVRs and 80 classes for CVRs.

In addition to the rarefaction simulations just described, we also used a rarefaction approach to test whether DVR models that are fit with high class numbers – and hence low sample sizes per class – perform worse than CVR models fit to the same resampled data. For these analyses, we used the same 20 class stratified bootstrap described in the last paragraph to generate 200 samples. For each species, we used one resampled data set size, between 1/4th and 1/16th of the original sample size; these sample sizes allowed us to fit a range of class numbers for each species but also resulted in small minimum sample sizes per class. We then fit models using from 5 up to 50 size classes, defining the upper and lower classes using the 1/class number and 1 – 1/class number quantiles and evenly divided class boundaries in between. We also fit CVR models to the same data and class boundaries, and employing median-based and CDF difference methods for discretization. We fit models to any data set that had at least 1 individual in each class, and also recorded the fraction of failed data sets for each size class number. We show results for class numbers for which < 30% of samples failed for DVR models. We then predicted lambda, damping ratio, and life span estimates for each model.

All analyses were conducting using R version 3.5.2 (R Core Team 2018). Example R scripts and data files showing the routines used in our analyses are included in Data S2.

Results

Use of mesh points vs CDF differences to characterize growth rates. To restrict the range of modeling decisions considered in subsequent analyses, we began by addressing the last modeling decision listed above (*Discretizing continuous size distributions*), asking if using mesh points or CDF differences had substantial effects on growth rate estimation and hence on model results. It is clear that the mesh point approach will be inaccurate at smaller class numbers (Fig. 2E), as it is essentially a crude numerical integration, but how accurate it is with larger class numbers under realistic assumptions is less clear. We therefore started by running a simplified simulation, not tied to any of our real data sets, to illustrate how well the mesh point method works to characterize growth when starting from a single size, and when the resulting sizes are far from size boundaries (so "eviction" is not a problem). As noted in Ellner et al. (2016; sections 2.7.4 and 6.8), the mesh point approach performs worst when there is low variance, in which case they suggest increasing the variance (if model predictions are unaffected), using sparse matrix methods with many mesh points, or alternative integration approaches such as Gauss-Legendre quadrature. To mimic this low-variance scenario, we simulated a realistically low SD of 0.5 with a mean size that varied from 49-51 (e.g., Figure 7A with mean = 50). We used from 10 to 100 mesh points spanning a range of sizes between 0 and 100 to discretize the resulting probability density function (PDF).

The most fundamental problem in approximating growth probabilities is if they do not sum to one, as all surviving individuals should have a size at the next time step; values greater than one implicitly boost survival when used in a full demographic model, while summed growth rates below one implicitly reduce survival. In our simulations, models with from 20 to 100 classes estimate growth probabilities that sum to between $3.0\text{x}10^{-5}$ and 1.76; while there is a general trend to more accuracy with higher class number, both under-and over estimates still occur as class number increases (Figure 7B). The summed growth probability is also highly sensitive to small differences in the mean of the growth distribution relative to the mesh points, with significant over- and under-estimation of total growth probability until high class numbers are reached (near to 100). This is because the probability of growth into a given class is estimated using exact PDF values at mesh points, so that the exact placement of mesh points relative to the peak of the PDF creates erratic mis-estimation, particularly when there are few mesh points relative to the growth variance. In contrast, the CDF difference approach always estimates the summed growth probability as one.

The problems with the mesh point approach can also affect the construction and results of full population models. To test the effects of mesh point mis-estimation on growth rate estimates across starting sizes, we built CVR models for each of our five focal species, using each of the two methods and both moderately high (50) and high (100) class numbers. For both methods, we normalized the estimated growth rates for a starting size based on the difference in the CDF between the minimum and maximum sizes used in the model. This test is the one proposed by Williams et al. (2012) to detect eviction. While it is the correct estimate of "true" eviction (i.e., growth outside the range of sizes in the model), and corrects all the transition probabilities for the CDF difference method so that they sum to one, for the mesh point models it is not a perfect test or correction, since the summed growth probabilities can deviate substantially from one, even in the absence of any meaningful eviction.

Use of mesh point methods leads to substantial over- or under-estimation of summed growth rates for some small or large classes for four of our five species, and does so even with models built with 50 or 100 classes (Figure 8). Mis-estimation tends to occur where variance in growth is low (Appendix S1: Figures S3-7), and can occur for size classes that include abundant individuals (e.g., large *Borderea* and small gorgonians). We also tested for the effects of eviction correction in changing the estimated mean and variance in growth for different sizes. Not surprisingly, eviction correction can substantially shift both mean and variance estimates (Appendix S1: Figures S8-S12); while this is expected, it does suggest that alternative models for growth rates that minimize the eviction problem need to be developed and more widely employed (e.g., Peterson et al. 2019).

To test the effects of mesh point mis-estimation of growth rates on lambda estimates, we took the same approach just described, but building entire demographic models for a range of class numbers for each focal species. The models used mean class values for estimation and even size class boundaries. For all of our focal species, use of mesh points results in greater deviations in lambda estimates and slower convergence on stable lambda values as class number increases than do models built using CDF differences to estimate growth probabilities (Figure 9).

In sum, the mesh point method can be highly inaccurate, and much of this inaccuracy will also be undetected by the usual test employed for growth rate eviction. Given that the CDF difference method is highly robust and extremely fast (only taking an additional 1.8 seconds than the mesh point approach in a test with 10,000 classes on a standard laptop), it is not clear that there is any reason to continue to use the mesh point approach to estimate growth, especially as it can generate artifacts and extreme class number dependence in the absence of any benefit. While the problems with the mesh point approach can be easily solved by using increasing class numbers, this number can reach ridiculous levels (e.g., > 4,000; Zuidema et al. 2010, Needham et al. 2018) and require careful analysis to detect. In the simulations below, we always use the CDF difference approach in our CVR models.

Effects of modeling decisions on population growth estimates. All four of the remaining modeling decisions that we explored can also have substantial influence on estimated lambda values. However, the strength and patterns of these effects are not necessarily what are usually assumed by most population biologists. While we have no independent measure of the "right" answer for these real data sets, most models based on the same data set converge on almost exactly the same lambda (λ) estimates with moderate (for discrete parameter estimation) or high (for continuous vital rate functions) class numbers, and we assume that these values are reasonable approximations of reality.

Class number and vital rate estimation method (DVR vs. CVR) show significant interactions in their effects on λ, but do not indicate any clear advantage for the use of CVRs (Figure 10). Models using DVRs, corresponding to classic matrix models, converge on the same lambda values as do models using CVRs, corresponding to IPM models, especially when using even class boundaries. λ values also show convergence to a very narrow range of values (± 0.01) by ~10–20 classes, depending on the species. While these are higher class numbers than are typical in many matrix models, they are far below those generally used in IPMs (Figure 4).

Neither estimation method gives consistently better results with small class numbers. DVRs always mis-estimated λ when used with very few classes, but the use of CVRs resulted in over or under-estimation, often of greater magnitude, depending on the species and other aspects of model construction. This suggests that the number of classes has a greater impact on model performance than the method of parameter estimation, but that the extremely large matrices used in most IPMs are unnecessary to achieve model accuracy.

In general, bistorts, *Borderea*, and guppies showed fastest convergence on the same lambda values with increasing classes than did the other two species, and the first two species also showed better correspondence between the predictions of median-based CVR and DVR models. Unfortunately, there is not a simple difference in the life histories or size distributions of the species that appears to correspond to these different results (see Figure 6, Appendix S1: Figures S3-7 for size distributions and vital rates of the species).

Other components of model building also influenced λ estimates. First, defining class boundaries using even divisions generally yielded more consistent results across class numbers, regardless of other modeling decisions. This was most obvious for bistort, *Borderea* and *Vulpicida*, for which even class models show convergence to the same lambda values at lower class numbers than do models with proportional classes, which continued to show divergence out to 100 classes. Second, use of estimated median sizes with CVR models to characterize average vital rates often yielded more accurate λ estimates, particularly at lower class numbers, than did use of midpoint sizes. This was especially striking for *Vulpicida*, for which use of midpoint sizes substantially altered λ estimates even with 80–100 classes with proportional class sizes, with no sign of convergence with the other estimates. The somewhat poorer performance of models with either proportional classes or midpoints appears to be due to the same underlying cause: mischaracterizing average performance either by grouping together very different individuals and/or by a poorer approach to characterizing the average state of individuals within a class.

One striking aspect of these results is that the outputs of CVR-based models are more dependent on other decisions about model structure and estimation than seems to be the case for discretely estimated vital rate models (Figure 10). This result contrasts with the common assumption that use of CVRs will lead to more stable, and hence reliable, results. This finding also suggests that, if models are made with moderate numbers of classes, discretely estimated matrix analyses are likely to provide estimates of growth rates that are just as robust as those arising from IPMs.

Simulation results: Damping ratios, lifespans, and sensitivity analysis. Predictions of lifespans and damping ratios mirrored those for population growth rates, although with greater effects of several modeling decisions (Figures 11 and 12). First, models made with even class widths show weaker effects of other modeling decisions in their eventual convergence than models made with proportional class widths. Second, CVR models that used median sizes to characterize performance often yielded more stable values than did those that used the midpoint of a class. This was most evident for life span estimates, for which midpoint models gave highly divergent estimates for *Vulpicida* and, with proportional classes, *Borderea* as well. Third, using discrete vital rate estimation yielded the same results as did CVR models, and generally converged upon stable values more quickly with increasing class number. Overall, these results bolster the conclusion that discretely-estimated matrix models are no less representative of demographic patterns than are IPMs fit with CVRs and evaluated at midpoints, with IPMs of high (>80) classes and matrix models of quite moderate size (~10-20 classes) giving essentially identical results.

We also examined the dependence of sensitivity analyses of lambda to matrix elements on modeling approach, contrasting the results of a DVR-based model of 20 even size classes and a CVR model of 80 even classes, evaluated at midpoints. The contrasting models for each species showed very similar results (Figure 13). The elasticity values of matrix elements were strongly correlated (r = 0.95-0.97) with no evidence of systematic bias and there was a similarly close correspondence in sensitivity values (r = 0.94-0.99; Appendix S1: Figure S18).

Simulation results: sample size effects. One assumed advantage of CVRs that has often been advanced in the demographic literature is that they perform better when data are sparse (Easterling et al. 2000, Ellner and Rees 2006, Ramula et al. 2009, Zuidema et al. 2010, Merow et al. 2014), so we compared model outputs for 20 class CVR models (fit with median values) vs. 20-class DVR models, each fit to 200 bootstrapped data sets across a range of sample sizes (Figure 14). Regardless of the modeling approach used, the variance in lambda estimates increased with smaller samples. However, to our surprise, there was little consistent advantage of the CVR approach with increasingly rarified data. Even when using sample sizes in the low hundreds, both discrete and continuous approaches to parameter estimation yielded similarly variable predictions. This finding does not support the generality that continuous functions will better estimate vital rates and hence produce better model outputs with small sample sizes. This result likely reflects in part the model-selection process inherent to fitting continuous vital rate functions. With smaller sample sizes, different sets of data can yield support for alternative forms of the vital rate functions, such as linear vs. quadratic functions, thereby altering model predictions. In addition, outliers can exert effects on entire vital rate functions when using the CVR approach, again creating variance in predictions that appear to be equivalent in their effects to the randomness generated in the estimates coming from the DVR approach. While this result might change with even smaller sample sizes, our simulation results based on 300–400 individual transitions already show so much variation that the effects of the sampling variance in the data overwhelm any advantage of one modeling approach over the other. We ran the same comparisons using 80 class CVR models (Appendix S1: Figure S19) and also used non-stratified bootstrapped data sets with proportional class boundaries (Appendix S1: Figure S20); in both cases, we find qualitatively similar results to those seen in the main simulations. In the future, it would be illuminating to estimate the relative contributions of different processes, such as model selection, outliers, and size distributions, to the precision of both DVR and CVR model predictions and how these vary with sample size for each approach.

We also conducted a different test of sample size effects, focused on the effects of low sample sizes per class for DVR models. Here we are looking at the possibility that with higher class numbers, DVR models will be increasing unreliable, since they will have at least some classes with vital rates estimated from extremely small samples (down to n=1 in our simulations). Using relatively low total sample sizes (see Methods) we fit models with a range of class numbers and found surprisingly little evidence for an advantage of CVR over DVR models or of a disadvantage of DVR models with higher class numbers (Figure 15 and Appendix S1: Figures S21-25), even when multiple classes have extremely low sample sizes (e.g., N< 6; Appendix S1: Figures S21-25). At the lowest sample sizes, for guppies and gorgonians we do see that DVR models generated a bimodal distribution of lambdas, but the second, erroneous peak in estimates results from models that have one or more classes where individuals are immortal and cannot leave, resulting in lambda =1. While erroneous, this is a pathology that is easy to recognize and rectify when building a model for a particular species. The surprising lack of advantage for CVR models also occurs for damping ratio and life span estimates (Appendix S1:

Figures S21-25), and does so even though we created models in which minimum sample sizes per class spanned extremely low numbers.

Simulation results: One more lesson. One lesson that we learned from constructing the models used in our simulations may not be apparent to many demographers, but can have large effects on model performance and especially the effects of varying class number: how the size distribution of new recruits is treated. For most species, including those in our focal data sets, there is a wide enough range of new recruit sizes, at least after their first year of life as a seedling or equivalent new recruit class, that they can grow into a meaningful range of sizes. While in IPM models, the size distribution of new recruits is typically quantified (92% of IPMs in our literature review; Data S1: Literature Review), in matrix models new recruits are often deposited into the smallest size class and then can proceed through the other size classes of a model. In the course of making our simulations, we realized that making this simple assumption guarantees an artificial dependency of model predictions on class number, because use of fewer, wider size classes essentially increases the size of new recruits, while narrower classes essentially shrink them. This problem is avoidable if, instead, new recruits are explicitly modeled as having probabilities of growing to a range of sizes. When testing effects of class number on model outputs, attention to this potential artifact is important.

DISCUSSION

IPMs and traditional matrix models are often discussed as wholly distinct modeling approaches, with IPMs representing a substantial improvement in demographic modeling by dealing more realistically with the continuous ranges of state variables and vital rates seen for many organisms. While an argument can be made that conceptually the two methods really are distinct, both our literature review and demographic analyses, based on data for five diverse organisms, challenge the view that they are entirely distinct in practice or that one is clearly superior. Instead, we find broad overlap in the way IPMs and matrix models are fit and interpreted. A quarter of size-based matrix models estimated at least one vital rate as a continuous function of size, and although these models used substantially fewer classes on average than IPMs, the range of class numbers was very wide and overlapping across the two approaches. Further, our simulations showed no substantive differences in outputs of models using discrete vs. continuous approaches to estimating vital rates. Rather, both approaches performed similarly well when sample sizes and class numbers were sufficient and similarly poorly when data were limiting or too few classes were used to capture an organism's life history. Further, we found little advantage to using more than 10-20 classes even for extremely slow-growing and long-lived organisms, suggesting that continuous demographic processes can be well approximated by matrices of moderate dimension. This range of size classes is at the high end for most size-based matrix models in the literature, but it is far lower than that used to analyze virtually all IPMs. Although these two aspects of demographic modeling - matrix dimension and discrete vs. continuous vital rate estimation - have received the most attention in the literature, our simulations also highlight the equal or greater importance of other modeling decisions, such as how classes are defined and continuous vital rate functions are discretized, as well as the quality and quantity of the underlying demographic data. Together, these results suggest that some model building decisions have been over-emphasized whereas data collection methods and sample size effects have been under-emphasized in discussions of improving demographic models and their predictions.

One of the principal critiques of traditional matrix models is that they use too few classes to accurately represent what are inherently continuous demographic processes, and this idea has been bolstered by analyses of IPMs that show that class numbers into the hundreds are often necessary to stabilize model outputs. Our simulations lend partial support to this idea, by showing that demographic models with too few classes do indeed produce biased outputs. Interestingly, however, our models were able to accurately capture the demography of long-lived species with sizes spanning up to 2.6 orders of magnitude with 10-20 classes – much less than what is typically used by IPMs. These results suggest that traditional matrix models for size-based life-histories may indeed require more classes than are typically used, although many models are built for species with smaller size ranges, more stage-based life histories, or shorter life spans than most of our focal species. For such species, there are likely to be smaller differences between the fates of most individuals and also less steep changes in vital rates across the sizes of most individuals. In such cases, fewer classes may well be sufficient, as we see with our guppy example. In addition, studies focused on particular species often make careful, biologically-based decisions about class boundaries which seem likely to yield better results than our more standardized but mindless class divisions (but see Ramula et al. 2020). Our results also suggest that the perceived need for extremely high class numbers with IPMs may be driven by the inefficiency of the midpoint method for estimating growth, or some other pathology in the way the models are being constructed, rather than a more fundamental need for high class number to capture biological patterns.

The other aspect of model fitting that has been emphasized in discussions of demographic modeling has been whether vital rates are estimated discretely for each class or by using the data across all classes to fit continuous vital rate functions (CVRs). The main arguments for CVRs are that they are more biologically realistic than discrete classes, are more accurate when data are limiting, can allow easier incorporation and testing of demographic drivers, and can utilize sophisticated statistical methods, such as mixed models or Bayesian approaches (Ramula et al. 2009, Merow et al. 2014b, Ehrlén et al. 2016, Elderd and Miller 2016). In our simulations, we found no evidence for the first two arguments. There was no improvement when using CVRs for a range of model outputs, including population growth, lifespan, damping ratio, or sensitivity and elasticity patterns. In fact, we observed a general tendency in our simulations for CVR-based model outputs to be more sensitive to other modeling decisions, such as whether classes are equally spaced or proportional to sample sizes. We also saw no evidence that CVRs increase the precision or accuracy of estimates as sample sizes decrease. This is in contrast to results found by Ramula et al. (2009), which compared 100-class IPMs with 4-6 class matrix models. This discrepancy may have been driven by the difference in class number rather than the method of vital rate estimation. By separating these two components in our simulations, we find a large effect of class number but little consistent effect of estimation method. Ramula et al. (2009) also found a weak correlation between matrix dimension and sample size for 63 plant matrix models, suggesting that matrix models with low sample sizes could also suffer from few classes. However, we found no relationship between matrix dimension and sample size for size-based matrix models in our literature review (Appendix S1: Fig. S2).

Our results lead us to conclude that neither the continuous nor discrete approaches to estimating vital rates is inherently better. Instead, the choice of approach should depend on the particular life history and analysis goals of a given study. Discrete vital rate estimation (DVR) is arguably the most flexible approach if life history patterns are complex, because it is agnostic about many aspects of vital rate patterns. For example, multiple state variables can be easily combined into complex states representing combinations of size, age, or stage variables. DVR can also easily accommodate sharp transitions or non-linearities in vital rates, cases where survival asymptotes at values less than 1, and cases that violate distributional assumptions about state variables (e.g., normally distributed growth). Further, the explanatory power of different model structures can be tested statistically to infer the number and placement of class divisions, and there is some evidence that this approach outperforms model structures informed by expert opinion alone (Ramula et al. 2020). Alternatively, there are clear advantages of the statistical framework of CVRs. By modeling vital rates in a regression-based framework, CVRs can easily incorporate the effects of covariates such as climate, can incorporate or correct for site or year effects as random variables, and can account for individual effects using random effects to account for repeated measures. CVRs can also provide a clear conceptual framework for hypothesis testing and model selection, and also allow investigation of sensitivities with respect to underlying functional forms or model parameters. However, it is worth noting that various methods, including multi-state mark recapture models, can also allow model selection to be applied to DVR estimation. Thus, CVR and DVR-based models may each be most appropriate for different datasets and analysis goals.

One surprising result to emerge from this work is the importance of other aspects of model fitting that have received far less attention in the demographic literature. For example, traditional matrix models often define classes based in part on sample sizes, but our results suggest that dividing classes evenly – on the transformed or untransformed scale for which size best relates to vital rates, depending on species -- generally gives more accurate results. We also identified several ways to improve methods for discretizing continuous vital rate functions. First, we show that the mesh point method can badly misestimate growth probabilities when using CVR functions, but that this is solved by using the CDF difference method. Second, our results suggest that vital rates are better characterized by using the median rather than the midpoint of a class, as long as the distribution of sampled individuals represents the size distribution in the population. We expect that this distinction is behind the slower convergence of the *Vulpicida* models with increasing class numbers. This is the only data set we used where sampling was not roughly comprehensive, but stratified over sizes, making the distribution of sampled individuals a poor characterization of the population-wide size distribution.

Finally, the most overwhelming effect on model accuracy in our simulations was the sample size of the underlying demographic data, which is an indicator of the sampling precision of vital rates and their relationships with the state variable used. We found low precision in model outputs with sample sizes less than several hundred regardless of modeling approach, and, in these cases, CVRs did not solve the problems caused by low sample sizes. The lowest sample sizes we used (305–405) were typical of many published demographic studies, and considerably larger than used in some studies (Appendix S1: Figures S2, S26). However, our simulations pooled all demographic data to estimate a single transition matrix. This means that spatial and temporal variation in individual fates will be at least somewhat larger in our samples than would be expected in samples of a single population over a single transition, potentially inflating the variance in lambda estimates we see with small samples. Furthermore, in cases where data are collected across multiple time periods or locations and sample sizes are limiting, CVRs may enable a researcher to improve accuracy by fitting vital rate functions that borrow strength by including all the data while appropriately modeling its structure. Statistical models can also be used to estimate DVRs while accounting for random effects of site or time period (Altwegg et al. 2007, Morris et al. 2011, Ramula et al. 2020), and a mixed strategy of fitting CVRs with class-specific random effects could also be useful in some cases.

There are several other important aspects of demographic model construction that we do not consider here, but that have recently been explored and shown to be of real importance. Perhaps most critically, we do not address how the choice of state variable can influence model results. As Louthan and Doak (2018) show, measured state variables that are not closely correlated with an individual's "true state" can yield misleading model results due to errors in characterizing state. For example, perennial plant size is often characterized by measuring leaf or stem size, whereas demographic rates may in fact be driven more by below-ground energy stores. A second concern that we do not address is the distributional assumptions that are often made when modeling size transitions with CVRs. Most IPMs assume that growth is normally distributed, but this can bias model outputs if growth is asymmetric such as when growth or shrinkage is more likely (Peterson et al. 2019). In general, choices regarding the form of vital rate functions are likely to be just as important for inference as the issues investigated here. Several authors have repeatedly emphasized that IPM practitioners should carefully evaluate the goodness of model fits for vital rate functions and their influence on IPM outputs (Easterling et al. 2000, Rees et al. 2014, Ellner et al. 2016), and we echo those recommendations here.

Taken together, our results suggest that IPMs and matrix models are overly dichotomized in the literature. We suggest that this distinction is neither useful nor representative of the range of modeling decisions that underlie every structured demographic model. Many models use some combination of continuous and discrete vital rate estimation (e.g., when some individuals are described by continuous state variables and others [juveniles, seeds, etc.] by stage). Although the methods used to build these matrices will often be very similar, the language and notation used to describe models are often completely different based on whether authors decide to refer to a model as an IPM or matrix model. In our experience, the terminology and integral notation used to describe IPMs can be intimidating to many new students and to non-specialist consumers of demographic analyses, such as conservation managers, when in fact the vital rate models and discretization methods would be familiar if described differently. In addition, much of the language used to present IPMs obscures the fact that the continuous vital rate functions are discretized into projection matrices prior to analysis, making the actual model outputs or their correspondence to matrix models difficult to understand. As we note above, many IPM studies in our literature review do not report the discretization methods used at all.

We suggest that it is more informative to refer to both projection matrix models and IPMs as Structured Population Models more generally, in part to emphasize the need to break these labels down into the important details of vital rate estimation, the number of size classes, and the methods used to discretize CVRs. In particular, we emphasize that statements suggesting that IPMs avoid discretization, are more biologically realistic, or perform better at small sample sizes, are not supported by our findings.

SPECIFIC RECOMMENDATIONS

Below, we highlight several of the most important recommendations for constructing demographic models that have emerged from this work.

1. When using an inherently continuous state variable, test the sensitivity of results to class number (particularly when using few classes).
2. When using continuous vital rate functions (CVRs), report methods for discretization by including class number and integration method.
3. When using CVRs to model size transitions, use the CDF difference method or explicitly show that the use of the standard mesh point method is accurate.
4. Especially when using smaller class numbers, use population size distributions to base vital rate estimates on representative (mean or median) sizes.

5. Both small sample sizes (indicative of low precision in vital rate estimates) or very few classes can result in biased or imprecise model outputs, and this should be carefully considered when interpreting or using published models (e.g., meta-analysis, COMPADRE or COMADRE databases), especially as many matrix models have been built with fewer classes than what we would recommend. A caveat to this conclusion is that for short-lived species or species with a limited range of sizes, fewer classes may be sufficient.

SUMMARY

In summary, we do not find support for several common generalities and assumptions about demographic modeling methods, and we also expose some new considerations for the construction of accurate structured population models. However, our results are generally positive: widely repeated but untested assumptions about the dependence of demographic results on modeling approaches were largely unsupported, meaning that we have a far wider range of useful demographic studies to learn from than would otherwise be the case. Looking forward, this result also implies that the structure and parameterization of demographic models should always be guided by careful consideration of the species and data being modeled and that, if this is done, different approaches will generally reach the same ecological conclusions. With sufficient sample sizes and enough classes to accurately represent the key life-history variation of a given species, a range of model-fitting approaches will converge on the same answer. Structured demographic models, one of the backbone methods of population biology, are comprised of a robust set of methods that can be usefully added to, but do not require fundamental re-tooling.

ACKNOWLEDGMENTS

Bruce Kendall provided many suggestions and critiques that greatly improved this work and its presentation. We also thank the many volunteers, colleagues, and assistants who helped gather the extensive data sets that we use in our simulation studies. Support for this work was provided by NSF awards 1146489, 1242558, 1242355, 1353781, 1340024, 1753980, and 1753954, 1144807, 0841423, and 1144083. Support also came from: USDA NIFA Postdoctoral Fellowship (award no. 2019-67012-29726/project accession no. 1019364) for RKS; the Swiss

Polar Institute of Food and Agriculture for NIC; the ICREA under the ICREA Academia Programme for CL; and SERDP contract RC-2512 and USDA National Institute of Food and Agriculture, Hatch project 1016746 for AML.

SUPPORTING INFORMATION

Additional supporting information may be found online at: [link to be added in production]

DATA AVAILABILITY

Example R scripts and data files showing the routines used in our analyses are included in Data S2.

LITERATURE CITED

Altwegg, R., M. Schaub, and A. Roulin. 2007. Age-specific fitness components and their temporal variation in the barn owl. The American Naturalist 169:47-61.

Bakker, V. J., and D. F. Doak. 2009. Population viability management: ecological standards to guide adaptive management for rare species. Frontiers in Ecology and the Environment 7:158-165.

Bakker, V. J., D. Doak, G. W. Roemer, D. K. Garcelon, T. J. Coonan, S. A. Morrison, C. Lynch, K. Ralls, and R. Shaw. 2009. Incorporating ecological drivers and uncertainty into a demographic population viability analysis for the island fox. Ecological Monographs 79:77-108.

Bakker, V., M. Finkelstein, D. F. Doak, L. Young, E. VanderWerf, J. Arata, P. Sievert, C. Vanderlip. 2018. The albatross of assessing and managing risk for long-lived pelagic yet terrestrial breeding species. Biological Conservation 217:83-95.

Barlow, J. and P. Boveng. 1991. Modeling age-specific mortality for marine mammal populations. Publications, Agencies and Staff of the U.S. Department of Commerce. 148.

Batista, W. B., W. J. Platt, and R. E. Machhiavelli. 1998. Demography of a shade-tolerant tree (*Fagus grandifolia*) in a hurricane-disturbed forest. Ecology 79:38-53.

Bernal, R. 1998. Demography of the vegetable ivory palm *Phytelephas seemannii* in Colombia, and the impact of seed harvesting. Journal of Applied Ecology 35:64-74.

Biek, R., W. C. Funk, B. A. Maxell, and L. S. Mills. 2002. What is missing in amphibian decline research: insights from ecological sensitivity analysis. Conservation Biology 16:728-734.

Carroll, R., C. Augspurger, A. Dobson, J. Franklin, G. Orians, W. Reid, R. Tracy, D. Wilcove, and J. Wilson. 1996. Strengthening the use of science in achieving the goals of the Endangered Species Act: an assessment by the Ecological Society of America. Ecological Applications 6:1-11.

Caswell, H. 2001. Matrix population models: construction, analysis, and interpretation. Sinauer Associates, Sunderland, Massachusetts.

Chien, P. D., P. A. Zuidema, and N. H. Nghia. 2008. Conservation prospects for threatened Vietnamese tree species: results from a demographic study. Population Ecology 50:227-237.

Cole, L. C. 1954. The population consequences of life history phenomena. The Quarterly Review of Biology 29:103-137.

Coma, R., M. Zabala, and J. M. Gili. 1995. Sexual reproductive effort in the Mediterranean gorgonian Paramuricea clavata. Marine Ecology Progress Series 117:173–183.

Crouse, D. T., L. B. Crowder, and H. Caswell. 1987. A stage-based population model for Loggerhead sea turtles and implications for conservation. Ecology 68:1412-1423.

Csergo, A. M., R. Salguero-Gomez, O. Broennimann, S. R. Coutts, A. Guisan, A. L. Angert, E. Welk, I. Stott, B. J. Enquist, B. McGill, J. C. Svenning, C. Violle, and Y. M. Buckley. 2017. Less favourable climates constrain demographic strategies in plants. Ecology Letters 20:969-980.

Dahlgren, J. P., and J. Ehrlén. 2009. Linking environmental variation to population dynamics of a forest herb. Journal of Ecology 97:666-674.

Dahlgren, J. P., M. B. Garcia, and J. Ehrlén. 2011. Nonlinear relationships between vital rates and state variables in demographic models. Ecology 92:1181-1187.

Dibner, R. R., M. L. Peterson, A. Louthan, and D. Doak. 2019. Multiple mechanisms confer stability to isolated populations of a rare endemic plant. Ecological Monographs 89:e01360.

Diez, J. M., I. Giladi, R. Warren, and H. R. Pulliam. 2014. Probabilistic and spatially variable niches inferred from demography. Journal of Ecology 102:544-554.

Doak, D. F., and W. F. Morris. 2010. Demographic compensation and tipping points in climate-induced range shifts. Nature 467:959-962.

Doak, D. F., G. K. Himes Boor, V. J. Bakker, W. F. Morris, A. Louthan, S. A. Morrison, A. Stanley, and L. B. Crowder. 2015. Recommendations for improving recovery criteria under the US Endangered Species Act. Bioscience 65:189-199.

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

Eberhardt. L.L. 1985. Assessing the Dynamics of Wild Populations. The Journal of Wildlife Management 49:997-1012

Ehrlén, J., W. F. Morris, T. von Euler, and J. P. Dahlgren. 2016. Advancing environmentally explicit structured population models of plants. Journal of Ecology 104:292-305.

Elderd, B., and T. E. 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.

Ellner, S. P., D. Z. Childs, and M. Rees. 2016. Data-driven modelling of structured populations: A practical guide to the integral projection model. Springer International Publishing, Switzerland.

Fitzpatrick S.W., G. S. Bradburd, C. T. Kremer, P. E. Salerno, L. M. Angeloni, and W. C. Funk. 2020. Genomic and fitness consequences of genetic rescue in wild populations. Current Biology 30:517-522.

Fitzpatrick S.W., J. C. Gerberich, L. M. Angeloni, L. L. Bailey, E. D. Broder, J. Torres-Dowdall, C. A. Handelsman, A. López-Sepulcre , D. N. Reznick, C. K. Ghalambor, and W. C. Funk . 2016. Gene flow from an adaptively divergent source causes rescue through genetic and demographic factors in two wild populations of Trinidadian guppies. Evolutionary Applications 9:879-891

Franco, M. and J. Silvertown. 2004. A comparative demography of plants based upon elasticities of vital rates. Ecology 85:531-538.

Franco, M., and J. Silvertown. 1996. Life history variation in plants: an exploration of the fast-slow continuum hypothesis. Philosophical Transactions of the Royal Society of London. B 351:1341-1348.

Gamelon, M., O. Gimenez, E. Baubet, T. Coulson, S. Tuljapurkar, and J. M. Gaillard. 2014. Influence of life-history tactics on transient dynamics: a comparative analysis across mammalian populations. American Naturalist 184:673-683.

Garcia, M. B. 2003. Demographic viability of a relict population of the critically endangered plant *Borderea chouardii*. Conservation Biology 17:1672–1680.

Gillespie, J. H. 1977. Natural selection for variances in offspring numbers: a new evolutionary principle. American Naturalist 111:1010-1014.

Griffith, A. B. 2017. Perturbation approaches for integral projection models. Oikos 126:1675-1686.

Gross, K., W. F. Morris, M. S. Wolosin, and D. F. Doak. 2005. Modeling vital rates improves estimation of population projection matrices. Population Ecology 48:79-89.

Haridas, C. V. and S. Tuljapurkar. 2005. Elasticities in variable environments: properties and implications. The American Naturalist 166:481-495

Hartshorn, G. S. 1975. A matrix model of tree population dynamics, p. 41-51. In F. B. Golley and E. Medina, eds. Tropical ecological systems: trends in terrestrial and aquatic research. Springer-Verlag, New York.

Hunter, C. M., H. Caswell, M. C. Runge, E. V. Regehr, S. C. Amstrup, and I. Stirling. 2010. Climate change threatens polar bear populations: a stochastic demography analysis. Ecology 91:2883-2897.

Jacquemyn, H., R. Brys, and E. Jongejans. 2010. Size-dependent flowering and costs of reproduction affect population dynamics in a tuberous perennial woodland orchid. Journal of Ecology 98:1204-1215.

Jongejans, E., O. Skarpaas, and K. Shea. 2008. Dispersal, demography and spatial population models for conservation and control management. Perspectives in Plant Ecology, Evolution, and Systematics 9:153-170.

Katz, D. S. 2016. The effects of invertebrate herbivores on plant population growth: a meta-regression analysis. Oecologia 182:43-53.

Kendall, B. E., M. Fujiwara, J. Diaz-Lopez, S. Schneider, J. Voigt, and S. Wiesner. 2019. Persistent problems in the construction of matrix population models. Ecological Modelling 406:33-43.

Lande, R. 1982. A quantitative genetic theory of life history evolution. Ecology 63:607-615.

Lande, R. 1988. Demographic models of the northern spotted owl (*Strix occidentalis caurina*). Oecologia 75:601-607.

Lefkovitch, L. P. 1965. The study of population growth in organisms grouped by stages. Biometrics 21:1-18.

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika 33:183-212.

Linares, C., and D. F. Doak. 2010. Forecasting the combined effects of disparate disturbances on the persistence of long-lived gorgonians: a case study of *Paramuricea clavata*. Marine Ecology Progress Series 402:59-68.

Linares, C., D. Doak, R. Coma, D. Díaz, and M. Zabala. 2007. Life history and viability of a long-lived marine invertebrate: the octocoral *Paramuricea clavata*. Ecology 88:918-928.

Louthan, A. M., R. M. Pringle, J. R. Goheen, T. M. Palmer, W. F. Morris, and D. F. Doak. 2018. Aridity weakens population-level effects of multiple species interactions on *Hibiscus meyeri*. Proceedings of the National Academy of Sciences 115:543-548.

Louthan, A., and D. Doak. 2018. Measurement error of state variables creates substantial bias in results of demographic population models. Ecology 99:2308-2317.

Matsuda, H., and K. Nichimori. 2003. A size-structured model for a stock-recovery program for an exploited endemic fisheries resource. Fisheries Research 60:223-236.

McEvoy, P. B., and E. M. Coombs. 1999. Biological control of plant invaders: regional patterns, field experiments, and structured population models. Ecological Applications 9:387-401.

Merow, C., A. M. Latimer, A. M. Wilson, S. M. McMahon, A. G. Rebelo, and J. A. Silander. 2014b. On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. Ecography 37:1167-1183.

Merow, C., J. P. Dahlgren, C. J. E. Metcalf, D. Z. Childs, M. E. K. Evans, E. Jongejans, S. Record, M. Rees, R. Salguero-Gómez, S. M. McMahon, and S. Ramula. 2014a. Advancing population ecology with integral projection models: a practical guide. Methods in Ecology and Evolution 5:99-110.

Metcalf, C. J. E., S. M. McMahon, R. Salguero-Gómez, E. Jongejans, and M. Rees. 2013. IPMpack: an R package for integral projection models. Methods in Ecology and Evolution 4:195-200.

Miller, T.E.X., and B. D. Inouye. 2011. Confronting two-sex demographic models with data. Ecology 92:2141–2151

Moloney, K. A. 1986. A generalized algorithm for determining category size. Oecologia 69:176-180.

Molowny-Horas, R., M. L. Suarez, and F. Lloret. 2017. Changes in the natural dynamics of *Nothofagus dombeyi* forests: population modeling with increasing drought frequencies. Ecosphere 8:1-17.

Montero-Serra, I., J. Garrabou, D. F. Doak, L. Figuerola, B. Hereu, J.-B. Ledoux, and C. Linares. 2018. Accounting for life-history strategies and timescales in marine restoration. Conservation Letters 11:e12341.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology. Sinauer, Sunderland, Massachusetts, USA.

Morris, W. F., J. Altmann, D. K. Brockman, M. Cords, L. M. Fedigan, A. E. Pusey, T. S. Stoinski, A. M. Bronikowski, S. C. Alberts, and K. B. Strier. 2011. Low demographic variability in wild primate populations: fitness impacts of variation, covariation, and serial correlation in vital rates. American Naturalist 177:E14-28.

Needham, J., C. Merow, C. H. Chang-Yang, H. Caswell, and S. M. McMahon. 2018. Inferring forest fate from demographic data: from vital rates to population dynamic models. Proceedings of the Royal Society of London B 285: 20172050.

Ozgul, A., T. Coulson, A. Reynolds, T. C. Cameron, and T. G. Benton. 2012. Population responses to perturbations: the importance of trait-based analysis illustrated through a microcosm experiment. American Naturalist 179:582-594.

Palmer, T. M., D. 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 107:17234-17239.

Peterson, M. L., W. Morris, C. Linares, and D. Doak. 2019. Improving structured population models with more realistic representations of non-normal growth. Methods in Ecology and Evolution 10:1431-1444.

Picard, N., and J. Liang. 2014. Matrix models for size-structured populations: unrealistic fast growth or simply diffusion? PlOS One 9:e98254.

Pulliam, H. R. 1988. Sources, sinks, and population regulation. American Naturalist 132:652-661.

R Core Development Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramula, S., and K. Lehtilä. 2005. Matrix dimensionality in demographic analyses of plants: when to use smaller matrices? Oikos 111:563-573.

Ramula, S., M. Rees, and Y. M. Buckley. 2009. Integral projection models perform better for small demographic data sets than matrix population models: a case study of two perennial herbs. Journal of Applied Ecology 46:1048-1053.

Ramula, S., N. Z. Kerr and E. E. Crone. 2020. Using statistics to design and estimate vital rates in matrix population models for a perennial herb. Population Ecology 62:53-63.

Rees, M., and S. P. Ellner. 2009. Integral projection models for populations in temporally varying environments. Ecological Monographs 79:575-594.

Rees, M., D. Z. Childs, and S. P. Ellner. 2014. Building integral projection models: a user's guide. Journal of Animal Ecology 83:528-545.

Rees, M., D. Z. Childs, J. C. Metcalf, K. E. Rose, A. W. Sheppard, and P. J. Grubb. 2006. Seed dormancy and delayed flowering in monocarpic plants: selective interactions in a stochastic environment. American Naturalist 168:E53-E71.

Rogers-Bennett, L., and D. W. Rogers. 2006. A semi-empirical growth estimation method for matrix models of endangered species. Ecological Modelling 195:237-246.

Sæther, B.-E., and Ø. Bakke. 2000. Avian life history variation and contribution of demographic traits to the population growth rate. Ecology 81:642-653.

Salguero-Gómez, R. and J. B. Plotkin. 2010. Matrix dimensions bias demographic inferences: implications for comparative plant demography. American Naturalist 176:710-722.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, C. Bein, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Romer, T. Ruoff, V. Sommer, J. Wille, J. Voigt, S. Zeh, D. Vieregg, Y. M. Buckley, J. Che-Castaldo, D. Hodgson, A. Scheuerlein, H. Caswell, and J. W. Vaupel. 2016a. COMADRE: a global data base of animal demography. Journal of Animal Ecology 85:371-384.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, A. Scheuerlein, D. A. Conde, E. Brinks, H. de Buhr, C. Farack, F. Gottschalk, A.

Hartmann, A. Henning, G. Hoppe, G. Römer, J. Runge, T. Ruoff, J. Wille, S. Zeh, R. Davison, D. Vieregg, A. Baudisch, R. Altwegg, F. Colchero, M. Dong, H. de Kroon, J.-D. Lebreton, C. J. E. Metcalf, M. M. Neel, I. M. Parker, T. Takada, T. Valverde, L. A. Vélez-Espino, G. M. Wardle, M. Franco, and J. W. Vaupel. 2015. The COMPADRE Plant Matrix Database: an open online repository for plant demography. Journal of Ecology 103:202-218.

Salguero-Gómez, R., O. R. Jones, E. Jongejans, S. P. Blomberg, D. Hodgson, C. Mbeau-Ache, P. A. Zuidema, H. De Kroon, and Y. M. Buckley. 2016b. Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. Proceedings of the National Academy of Sciences 113:230-235.

Schemske, D. W., B. C. Husband, M. H. Ruckelshaus, C. Goodwillie, I. M. Parker, and J. G. Bishop. 1994. Evaluating approaches to the conservation of rare and endangered plants. Ecology 75:584-606.

Sergei, I. 2015. binr: Cut Numeric Values into Evenly Distributed Groups. R package version 1.1. https://CRAN.R-project.org/package=binr

Shaffer, M. L. 1981. Minimum population sizes for species conservation. Bioscience 31:131-134.

Shaffer, M. L. 1983. Determining minimum viable population sizes for the grizzly bear. Bears: their biology and management 5:133-139.

Shea, K. 1998. Management of populations in conservation, harvesting and control. Trends in Ecology and Evolution 13:371-375.

Shimatani, I. K., Y. Kubota, K. Araki, S.-I. Aikawa, and T. Manabe. 2007. Matrix models using fine size classes and their application to the population dynamics of tree species: Bayesian non-parametric estimation. Plant Species Biology 22:175-190.

Shriver, R. K., C. Andrews, R. Arkle, D. Barnard, M. Duniway, M. J. Germino, D. S. Pilliod, D. Pyke, J. Welty, and J. Bradford. 2019. Transient population dynamics impede restoration and may promote ecosystem transformation after disturbance. Ecology Letters 22:1357-1366

Shriver, R. K., K. Cutler, and D. F. Doak. 2012. Comparative demography of an epiphytic lichen: support for general life history patterns and solutions to common problems in demographic parameter estimation. Oecologia 170:137-146.

Siler, W. 1977. A Competing-Risk Model for Animal Mortality. Ecology 60:750-757

Silvertown, J., M. Franco, I. Pisanty, and A. Mendoza. 1993. Comparative plant demography: relative importance of life-cycle components to the finite rate of increase in woody and herbaceous perennials. Journal of Ecology 81:465-476.

Stearns, S. C. 1992. The evolution of life histories. Oxford University Press, Oxford.

Stubben, C. J., and B. G. Milligan. 2007. Estimating and analyzing demographic models using the popbio package in R. Journal of Statistical Software 22:11.

Ureta, C., C. Martorell, J. Hortal, and J. Fornoni. 2012. Assessing extinction risks under the combined effects of climate change and human disturbance through the analysis of life-history plasticity. Perspectives in Plant Ecology, Evolution, and Systematics 14:393-401.

Vandermeer, J. 1978. Choosing category size in a stage projection matrix. Oecologia 32:79-84.

White, J. W., K. J. Nickols, D. Malone, M. H. Carr, R. M. Starr, F. Cordoleani, M. L. Baskett, 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.

Wilber, M. Q., R. A. Knapp, M. Toothman, and C. J. Briggs. 2017. Resistance, tolerance and environmental transmission dynamics determine host extinction risk in a load-dependent amphibian disease. Ecology Letters 20:1169-1181.

Williams, J. L. 2009. Flowering life-history strategies differ between the native and introduced ranges of a monocarpic perennial. American Naturalist 174:660-672.

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

Yau, A. J., H. S. Lenihan, and B. E. Kendall. 2014. Fishery management priorities vary with self-recruitment in sedentary marine populations. Ecological Applications 24:1490-1504.

Yokomizo, H., T. Takada, K. Fukaya, and J. G. Lambrinos. 2017. The influence of time since introduction on the population growth of introduced species and the consequences for management. Population Ecology 59:89-97.

Zuidema, P. A. 2000. Demography of exploited tree species in the Bolivian Amazon. Utrecht University, Utrecht.

Zuidema, P. A., E. Jongejans, P. D. Chien, H. J. During, and F. Schieving. 2010. Integral Projection Models for trees: a new parameterization method and a validation of model output. Journal of Ecology 98:345-355.

TABLES

Table 1: Summary of studies included in the literature review.

			Deterministic		Stochastic	
	Number models	Number species	prop. Matrix	prop. IPM	prop. Matrix	prop. IPM
Annual forbs	24	24	0.75	0.0	0.25	0
Perennial forbs	314	236	0.46	0.18	0.27	0.08
Woody plants	255	222	0.45	0.28	0.19	.08
Algae, lichen, & mosses	13	12	0.31	0	0.69	0
Other non-woody plants*	92	64	0.39	0.40	0.12	0.09
Total autotrophs	**698**	**558**	**0.46**	**0.24**	**0.22**	**0.08**
Amphibians	20	19	0.55	0.15	0.30	0.0
Reptiles	29	24	0.69	0.03	0.24	0.03
Fishes	63	45	0.62	0.22	0.13	0.03
Birds	134	108	0.68	<0.01	0.31	<0.01
Mammals	240	175	0.78	0.03	0.18	0.01
Total vertebrates	**486**	**371**	**0.71**	**0.06**	**0.22**	**0.01**
Invertebrates	**87**	**72**	**0.69**	**0.10**	**0.15**	**0.06**
Total	**1271**	**1001**	**0.57**	**0.16**	**0.22**	**0.05**

* Including ferns, graminoids, and Cactaceae

FIGURE LEGENDS

Figure 1: Diagram of the key steps and considerations in fitting either a matrix model or Integral Projection Model (IPM). Both approaches begin and end with the same steps (yellow boxes) but may differ in their methods of parameter estimation (blue boxes) and resulting matrix dimensions (green boxes).

Figure 2. Comparison of methods used to calculate probabilities of growing from a given starting size into a given size bin (x). A) The probability density of size at t+1 can be discretized into classes (defined by black lines) with midpoint sizes (circles). B) The probability of growing into a particular class is most often approximated by the midpoint method, by evaluating the probability density at the midpoint and multiplying by the class width (h). C) Alternatively, the probability of growing into a class is given exactly by the difference in the cumulative probability function (CDF) values at the bin edges. The approximation in B is accurate with many narrow bins relative to the variance in growth (D), but can be poor if classes are wide relative to the variance in growth (E). The actual growth probabilities based on differences in the cumulative density function or CDF (red dots) sum to 1, whereas the approximated growth probabilities based on point estimates from the PDF (blue dots) may be less than or greater than 1.

Figure 3: Changes in published demographic models over time. Circles are the proportion of A) models that are described as IPMs vs. matrix models, and B) matrix models that use continuous vital rate (CVR) estimation published between 2002 and 2018, with fitted relationships over time. Proportions are shown for all models (open circles, dashed lines) or only size-based models (filled circles, solid lines). Circle size is proportional to the total number of models.

Figure 4. Histograms of the number of classes used in demographic models published between 2002 and 2018. Histograms are shown for A) all models or B) models with size as the state variable, divided between matrix models (solid black lines) and Integral Projection Models (IPMs; dashed red lines). Histogram values are shown as points connected by a line.

Figure 5: The empirical work flow for demographic models published between 2002 and 2018. Arrow and circle widths indicate the number of corresponding models. Note that all Integral Projection Model (IPM) parameter estimation is classified as continuous.

Figure 6. Size distributions of individuals included the example data sets used for demographic simulations, shown on the scales used to construct the demographic models. Note that for all species other than *Borderea*, before using the data in simulations we set a maximum size that

eliminated long, very sparse tails of larger sizes, as these created problems for the automated binning of data used in our simulations.

Figure 7. An illustration of the problems with standard mesh point growth estimation. A) A distribution of ending size values, with mean 50 and SD 0.5, on a possible size range between 0 and 100 and with probabilities shown for classes of width 1. B) The summed probability of growing to any size, as estimated by the standard mesh point method. Results are shown for bin numbers between 10 and 100. Each line shows results for a different mean ending size between 49 and 51. Any reasonable way to estimate growth rates should yield a summed probability of exactly one. Unless high mesh point numbers are used, the mesh point method yields values much larger or smaller than one most of the time. With a narrower ending size distribution, far higher bin numbers are needed to yield reasonable results.

Figure 8. Even at high class numbers, the mesh point approach yields inaccurate growth rates for some starting size classes. Results show the summed growth probabilities for each starting size for models built with 50 (green) or 100 (blue) classes for each of our focal species. For four of the five species, the mesh point method gives poor total growth estimates for some large or small size classes, even with 100 classes.

Figure 9. Estimates of lambda derived from continuous vital rate (CVR) models built with either the standard mesh point approach to estimate growth or using the cumulative distribution function (CDF) difference approach, shown for each species. The mesh point approach yields much worse estimates of lambda at low to moderate class numbers.

Figure 10. Lambda values with varying class number and modeling approaches for each focal species. Models are built using class divisions that evenly divide the size range and have a constant width (left) or that proportionally divide the size range to include similar sample sizes within each class (right). Within each panel, data are shown for models built with discrete vital rate estimation (DVR) or with continuous vital rate function using either midpoint (CVR – midpoint) or median (CVR – median) size estimates of class-specific vital rates. Note the different range of class numbers on the x-axis for guppies.

Figure 11. Life span estimates for four species, with varying class number and modeling approaches. Models are built using class divisions that evenly divide the size range and have a constant width (left) or that proportionally divide the size range to include similar sample sizes within each class (right). Within each panel, data are shown for models built with discrete vital rate estimation (DVR) or with continuous vital rate function using either midpoint (CVR – midpoint) or median (CVR – median) size estimates of class-specific vital rates.

Figure 12. Damping ratios with varying class number and modeling decisions. Models are built using class divisions that evenly divide the size range and have a constant width (left) or that proportionally divide the size range to include similar sample sizes within each class (right). Within each panel, data are shown for models built with discrete vital rate estimation (DVR) or with continuous vital rate function using either midpoint (CVR – midpoint) or median (CVR – median) size estimates of class-specific vital rates.

Figure 13. Close correspondence between elasticity values generated by a 20 class DVR and an 80 class continuous vital rate (CVR) model for each species. To condense values from the larger model, we summed sets of elasticities in each column of the matrix that corresponded to a single, broader category in the smaller model, and then took the average of these values across columns corresponding to the categories in the smaller model. We only compare elasticities for non-zero matrix elements. Pearson correlation coefficients are shown for each relationship, and lines give the 1:1 slope. Appendix S1: Figure S18 shows a comparable figure for sensitivity values.

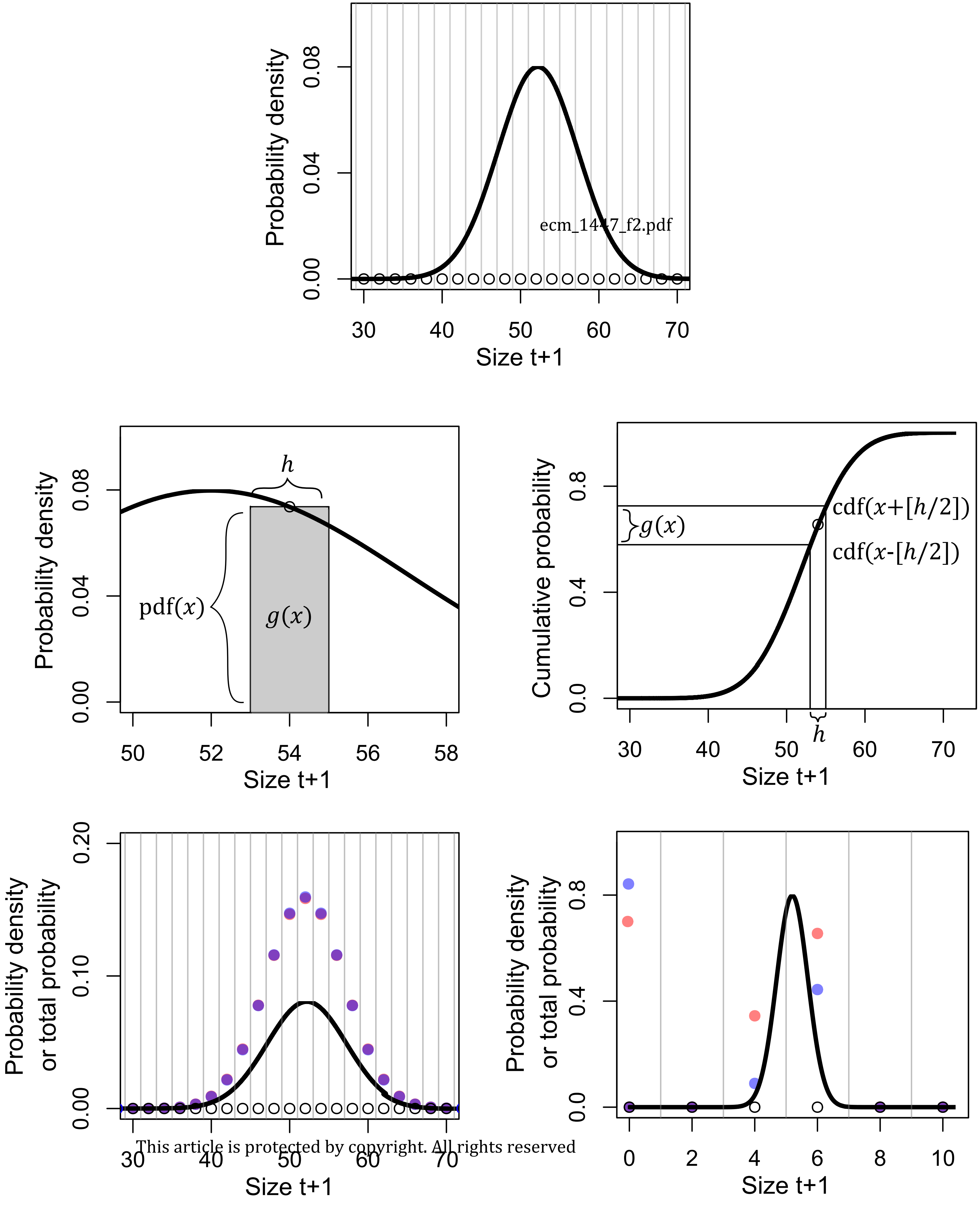
Figure 14. Distribution of lambda estimates with declining sample sizes. For each of 200 random draws of data with decreasing sample sizes, we fit a 20 bin continuous vital rate (CVR) based model (gray) or a 20 bin DVR-based model (black). CVR models were fit using median bin characterization and all models used even size classes (see Methods for more description of the size class boundaries). The horizontal line running through each distribution shows the mean. See Figure S19 for a comparison of the 20 class DVR models with 80 class CVR models.

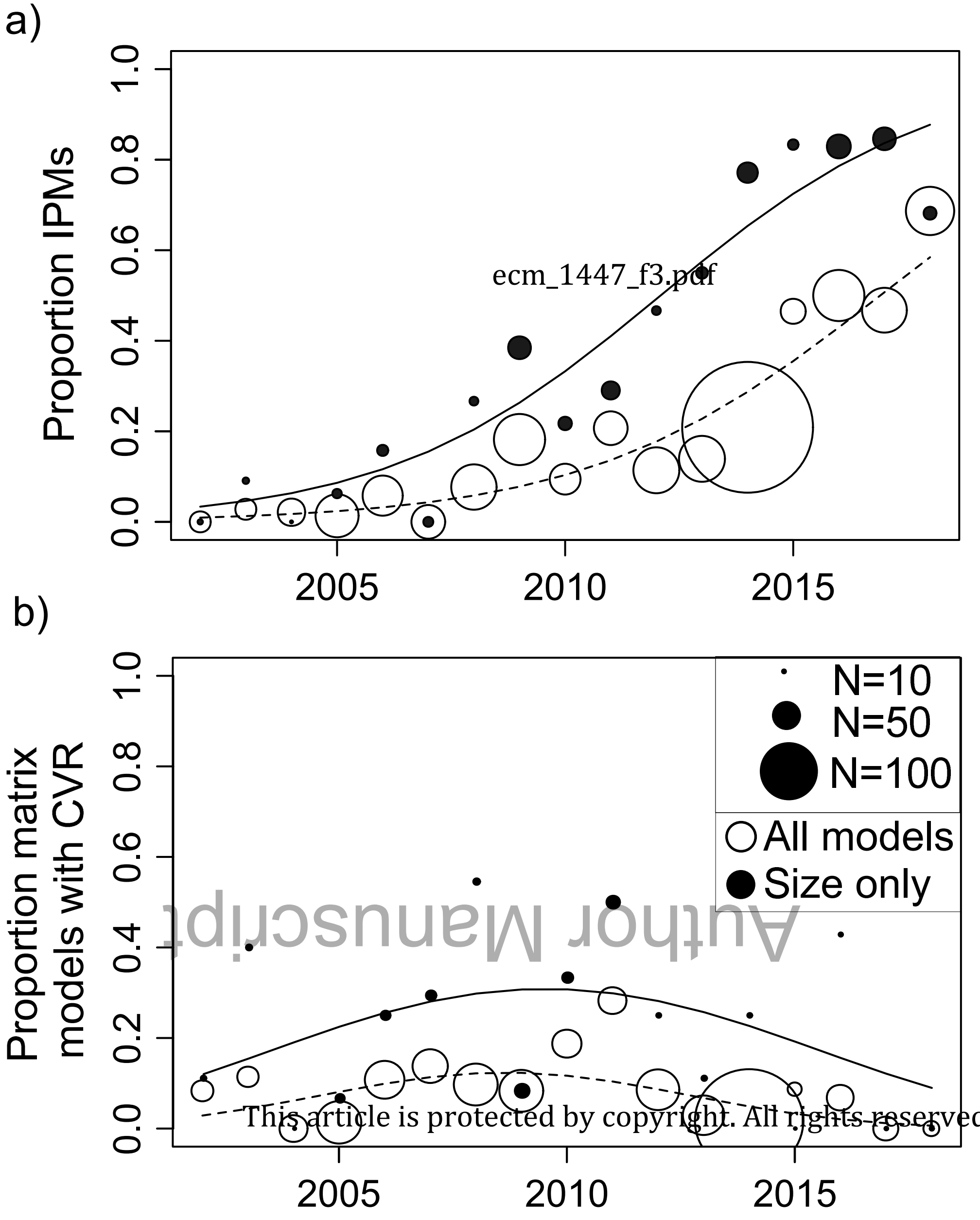
Figure 15. Distribution of lambda estimates with increasing size classes. For each of 200 random draws of data with a single decreased sample size, we fit a 20 class continuous vital rate (CVR) based model (gray) or a 20 class DVR-based model (black). CVR models were fit using median class characterization and all models used even size classes. The horizontal line running through each distribution shows the mean. See Figures S21-25 for comparable results for damping ratios and life spans and for summaries of size-class specific sample sizes.

1. Demographic data

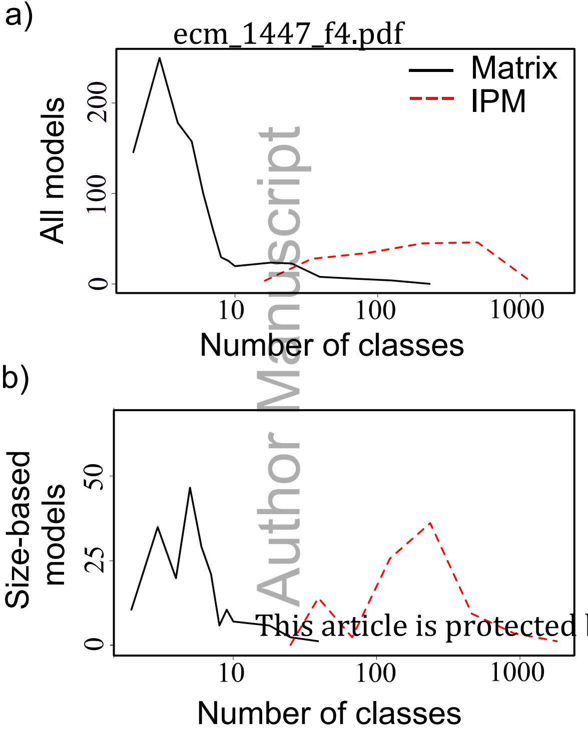
Vital rates (e.g., individual growth, survival, & reproduction)

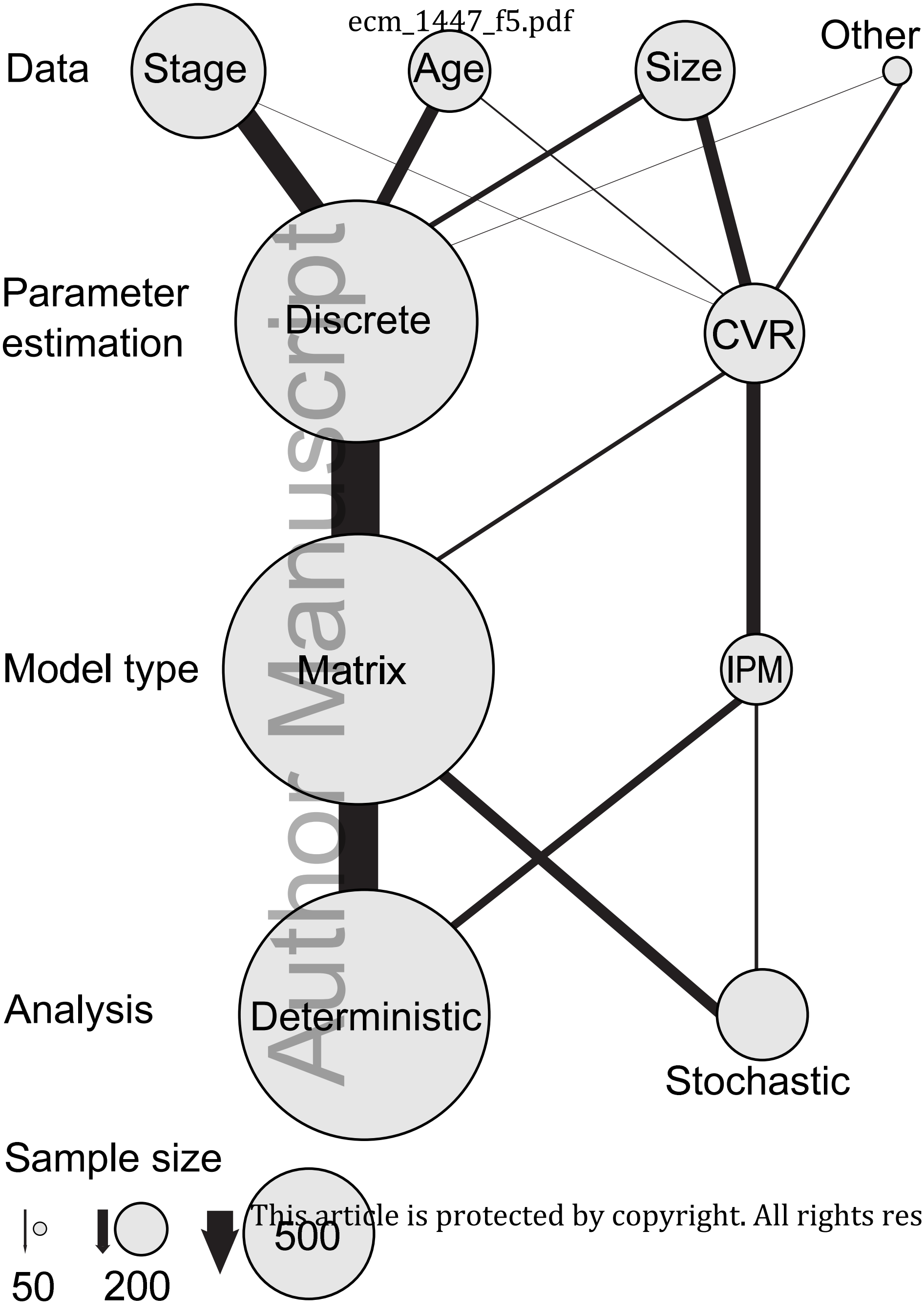
One or more state variables (e.g., size, age, life history stage)

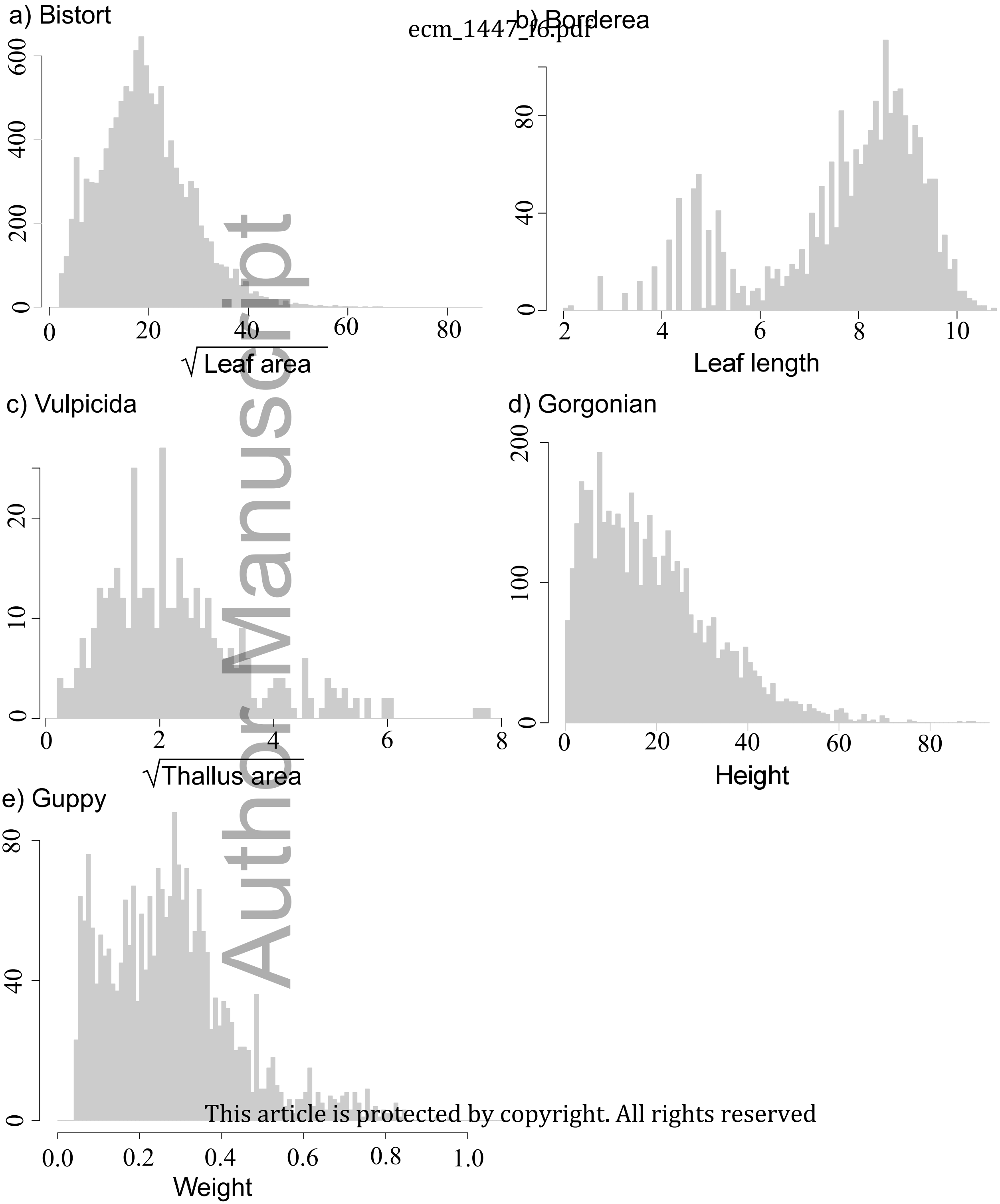

Probability density
0.00
0.04
0.08
ecm_1447_f2.pdf
30
40
50
60
70
Size t+1
Probability density
0.00
0.04
0.08
h
$\text{pdf}(x)$
$g(x)$
50
52
54
56
58
Size t+1
Cumulative probability
0.0
0.4
0.8
$g(x)$
$\text{cdf}(x+[h/2])$
$\text{cdf}(x-[h/2])$
h
30
40
50
60
70
Size t+1
Probability density
or total probability
0.00
0.10
0.20
30
40
50
60
70
Size t+1
Probability density
or total probability
0.0
0.4
0.8
0
2
4
6
8
10
Size t+1

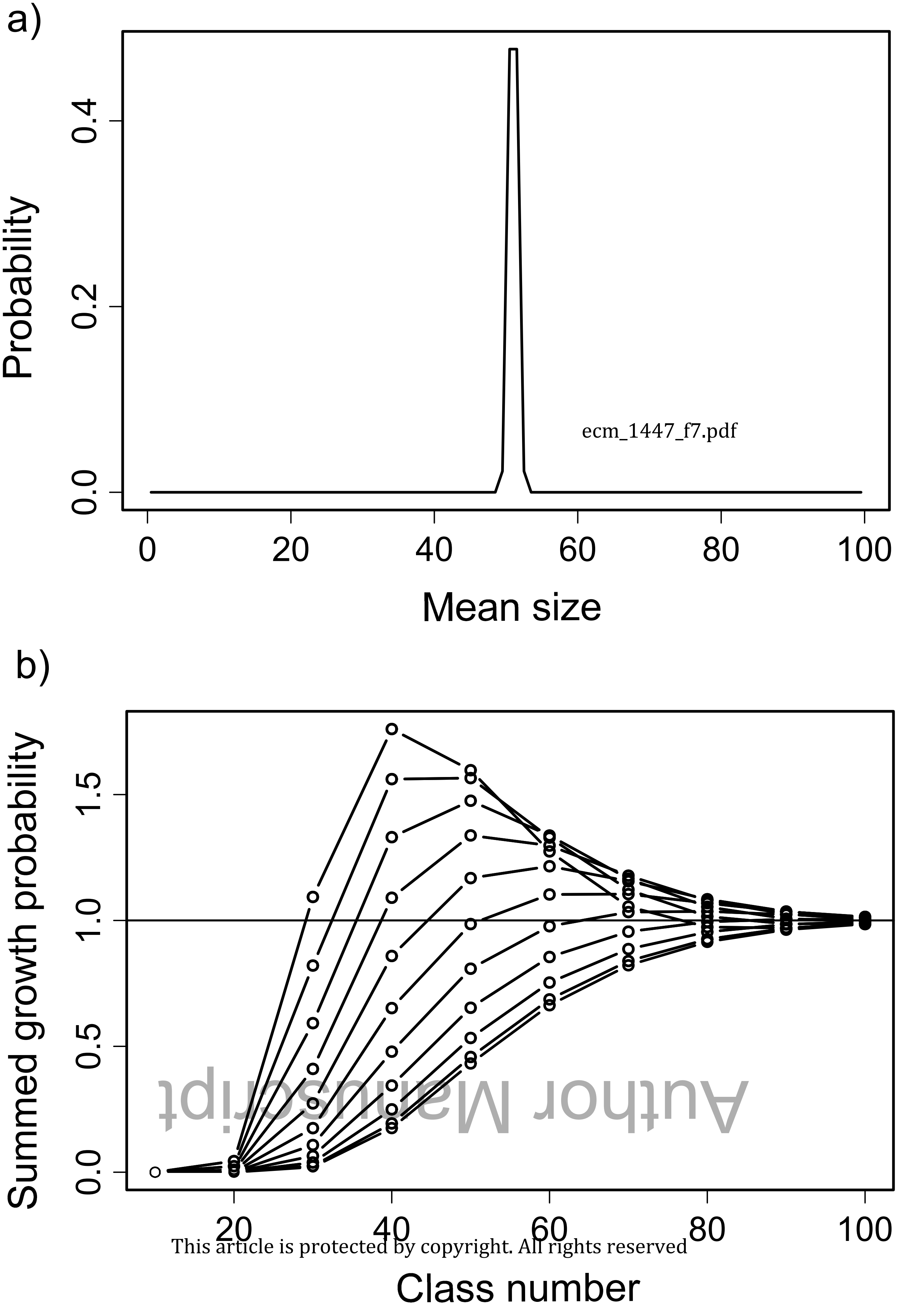
a)
Proportion IPMs
0.0
0.2
0.4
0.6
0.8
1.0
2005
2010
2015
b)
Proportion matrix models with CVR
0.0
0.2
0.4
0.6
0.8
1.0
2005
2010
2015
N=10
N=50
N=100
All models
Size only

ecm_1447_f4.pdf

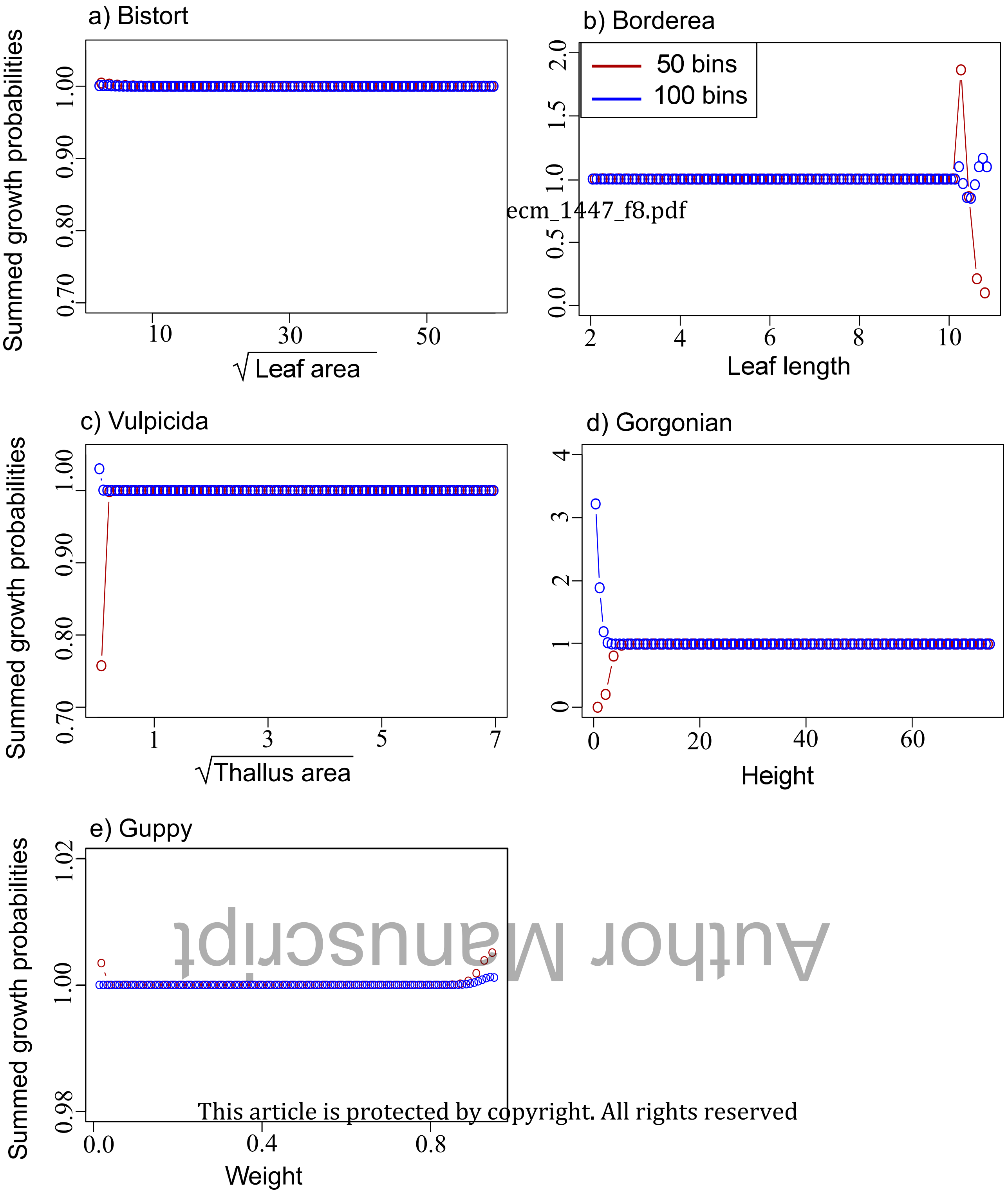


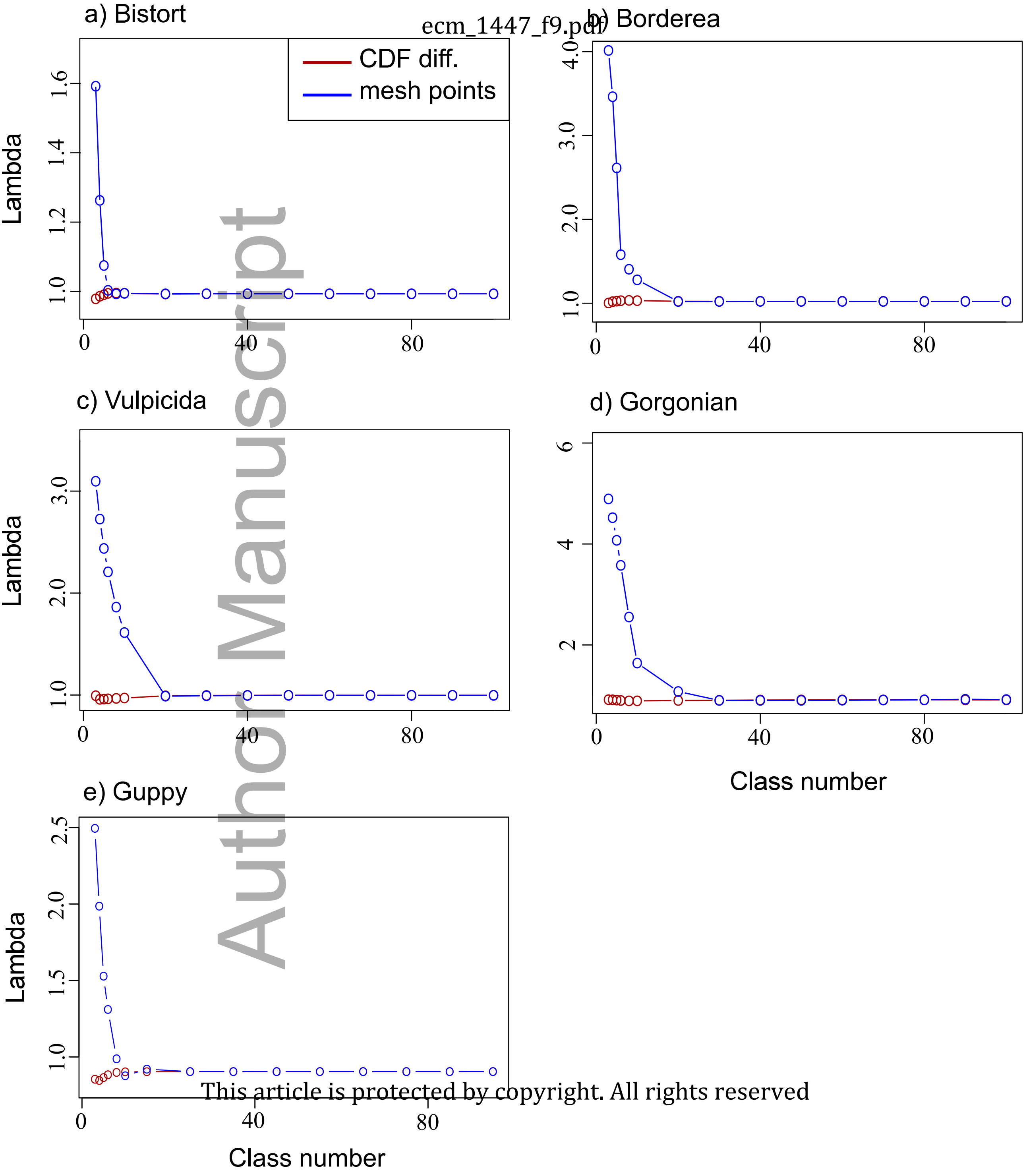

Data
Stage
Age
Size
Other
Parameter estimation
Discrete
CVR
Model type
Matrix
IPM
Analysis
Deterministic
Stochastic
Sample size
50
200
500

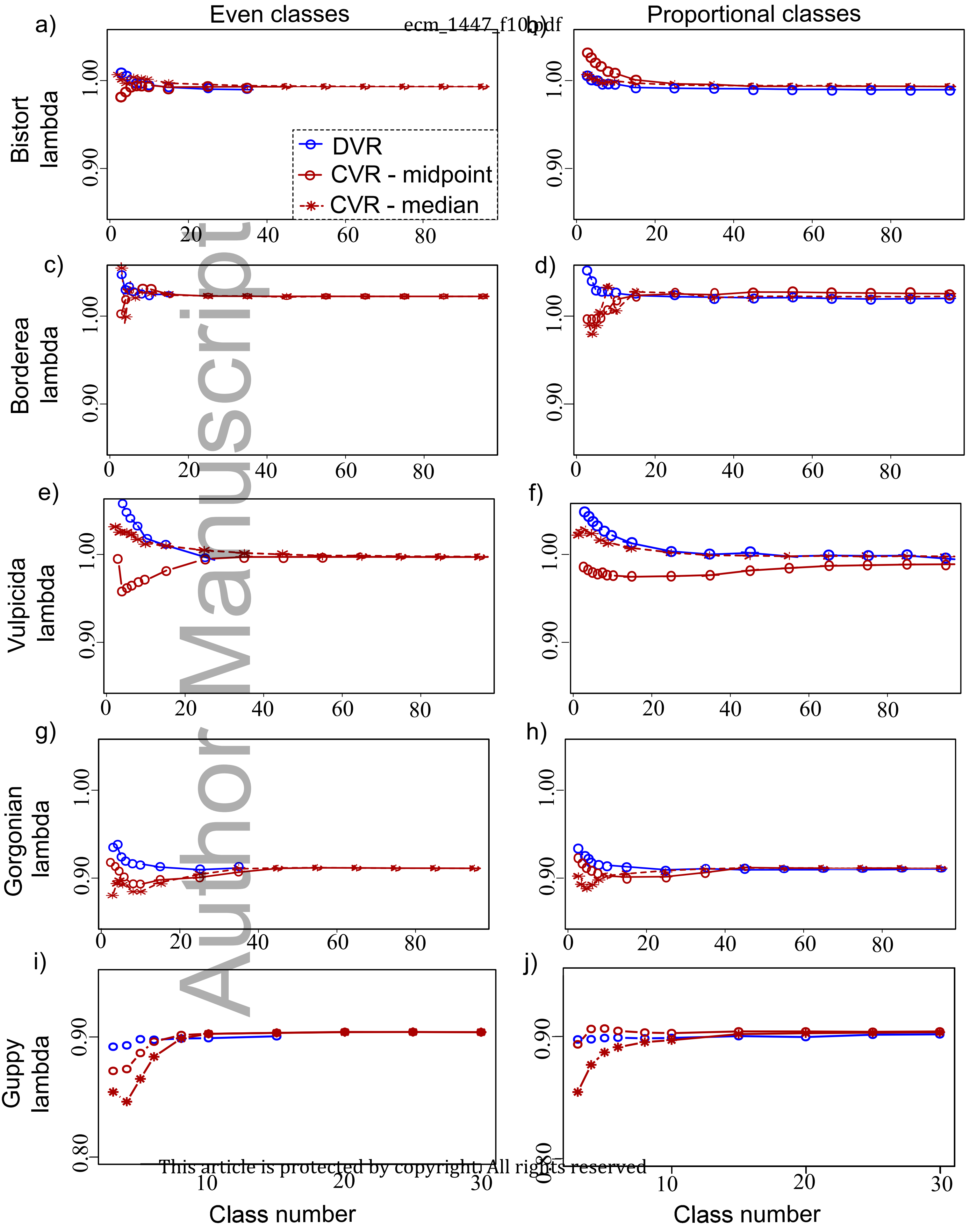

a) Bistort
0
200
400
600
0
20
40
60
80
√ Leaf area
b) Borderea
0
40
80
2
4
6
8
10
Leaf length
c) Vulpicida
0
10
20
0
2
4
6
8
√Thallus area
d) Gorgonian
0
100
200
0
20
40
60
80
Height
e) Guppy
0
40
80
0.0
0.2
0.4
0.6
0.8
1.0
Weight

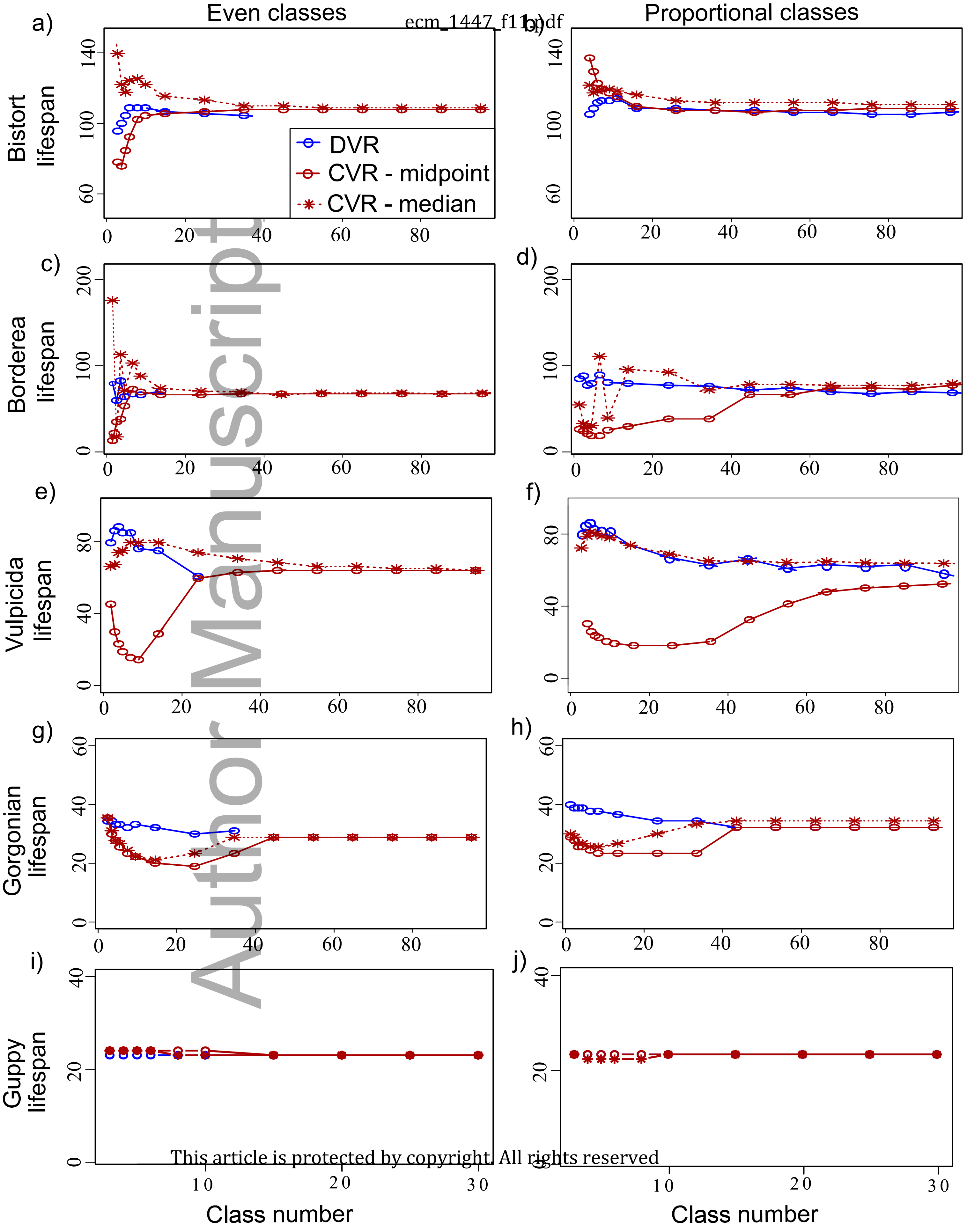
a)
Probability
0.0
0.2
0.4
0
20
40
60
80
100
Mean size
ecm_1447_f7.pdf
b)
Summed growth probability
0.0
0.5
1.0
1.5
20
40
60
80
100
Class number

ecm_1447_f8.pdf

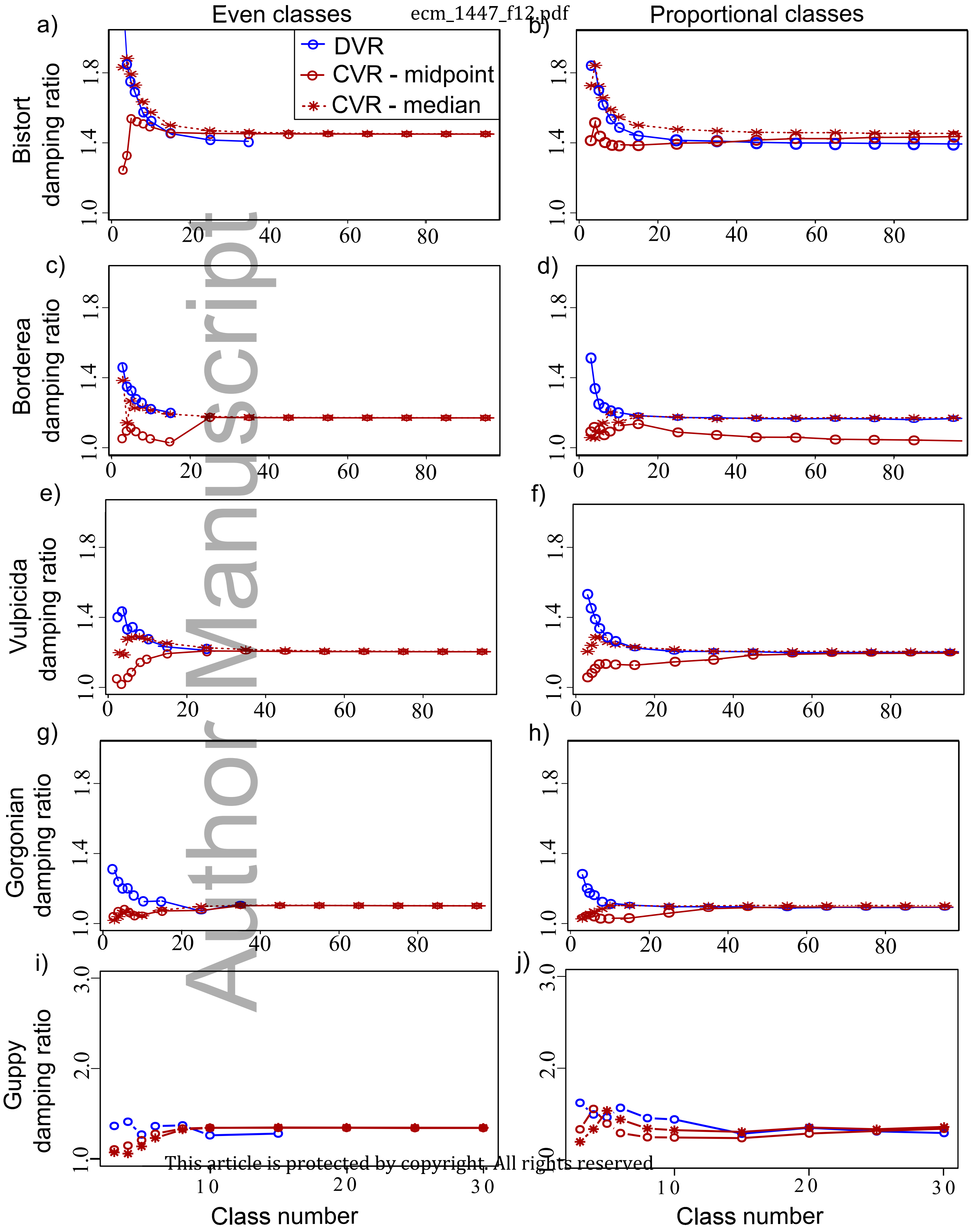



a) Bistort
b) Borderea
c) Vulpicida
d) Gorgonian
e) Guppy
CDF diff.
mesh points
Lambda
Class number
1.0
1.2
1.4
1.6
2.0
3.0
4.0
2
4
6
1.5
2.5
0
40
80

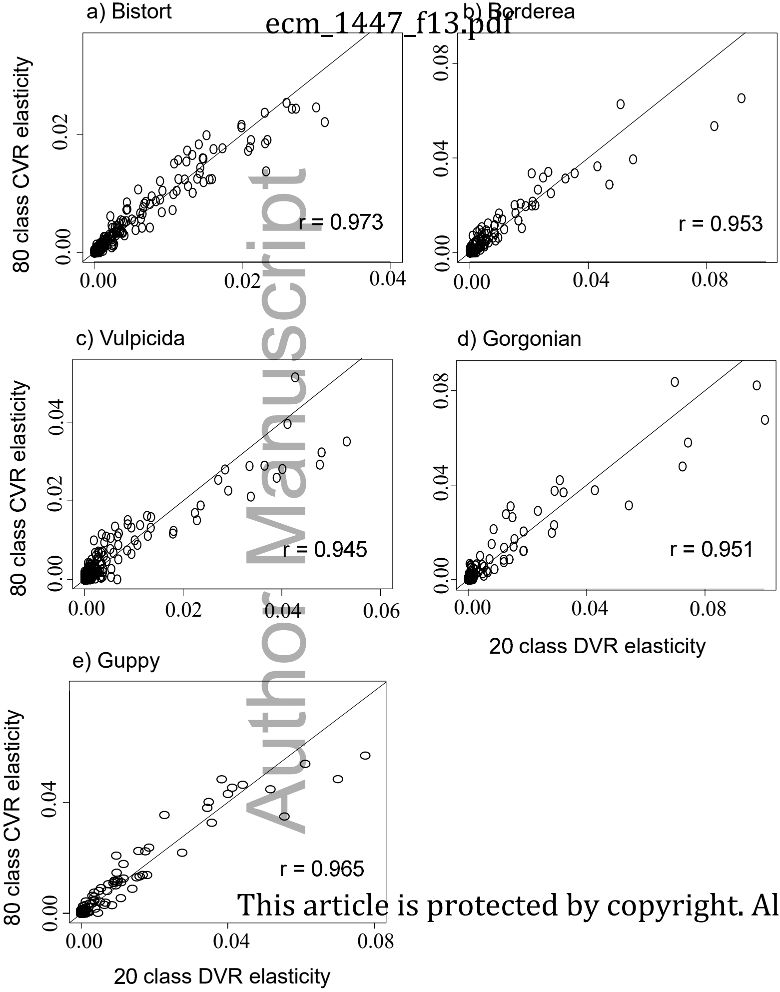

Even classes
Proportional classes
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
Bistort lambda
Borderea lambda
Vulpicida lambda
Gorgonian lambda
Guppy lambda
DVR
CVR - midpoint
CVR - median
0.90
1.00
0.80
0
20
40
60
80
10
30
Class number

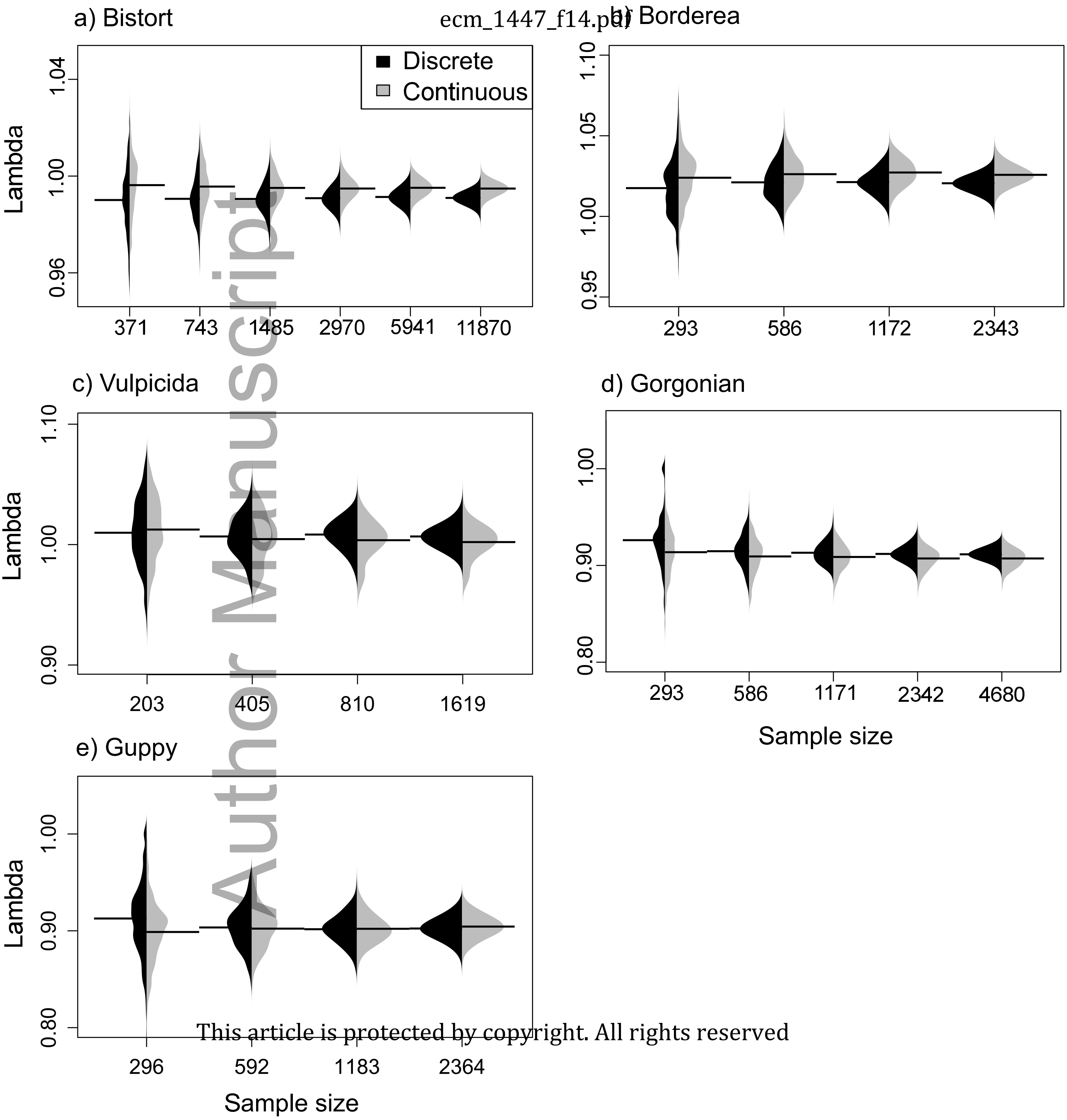

Even classes
Proportional classes
a)
b)
c)
d)
e)
f)
g)
h)
i)
j)
Bistort lifespan
Borderea lifespan
Vulpicida lifespan
Gorgonian lifespan
Guppy lifespan
DVR
CVR - midpoint
CVR - median
140
100
60
200
100
0
80
40
0
60
40
20
0
40
20
0
0
20
40
60
80
10
20
30
Class number
Class number

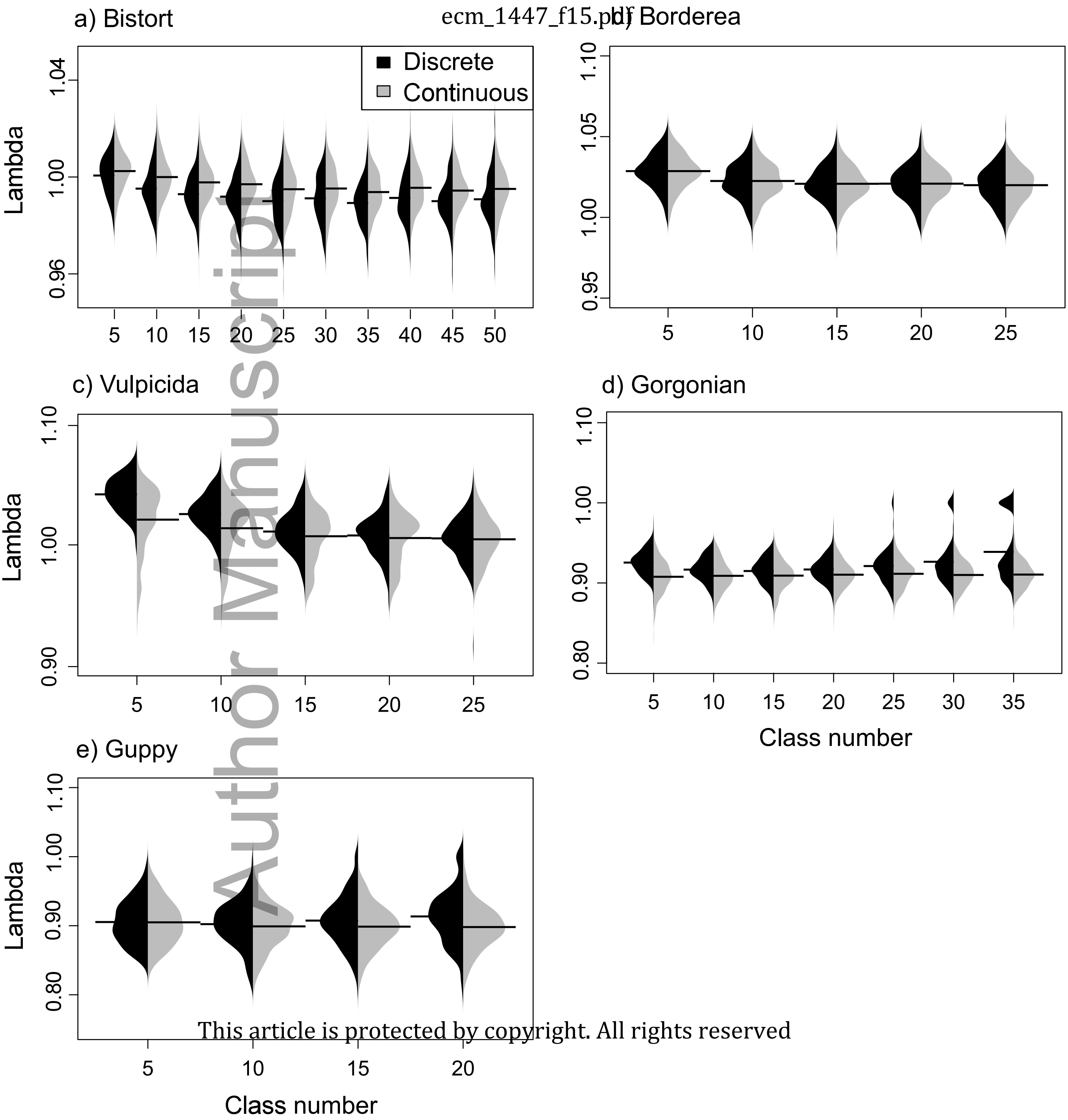
ecm_1447_f12.pdf



ecm_1447_f13.pdf




a) Bistort
Discrete
Continuous
Lambda
1.04
1.00
0.96
371
743
1485
2970
5941
11870
b) Borderea
1.10
1.05
1.00
0.95
293
586
1172
2343
c) Vulpicida
Lambda
1.10
1.00
0.90
203
405
810
1619
d) Gorgonian
1.00
0.90
0.80
293
586
1171
2342
4680
Sample size
e) Guppy
Lambda
1.00
0.90
0.80
296
592
1183
2364
Sample size


a) Bistort
■ Discrete
▨ Continuous
Lambda
1.04
1.00
0.96
5
10
15
20
25
30
35
40
45
50
b) Borderea
1.10
1.05
1.00
0.95
5
10
15
20
25
c) Vulpicida
Lambda
1.10
1.00
0.90
5
10
15
20
25
d) Gorgonian
1.10
1.00
0.90
0.80
5
10
15
20
25
30
35
Class number
e) Guppy
Lambda
1.10
1.00
0.90
0.80
5
10
15
20
Class number