Running title: Comment on Doak et al. (2021)

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

Stephen P. Ellner*[a],
Peter B. Adler[b], Dylan Z. Childs[c], Giles Hooker[d], Tom E.X. Miller[e] and Mark Rees[c]

[a]Department of Ecology and Evolutionary Biology,
Cornell University, Ithaca, New York 14853 USA
[b]Department of Wildland Resources and the Ecology Center,
Utah State University, Logan, Utah USA
[c]Department of Plant and Animal Sciences, University of Sheffield, Sheffield S10 2TN UK
[d]Department of Statistics and Data Science,
Cornell University, Ithaca, New York 14853 USA
[e]Department of BioSciences, Rice University, Houston, Texas 77005 USA

September 28, 2021

Final version

Keywords: Demography, modeling, statistics, structured population models

Open research statement: No data were collected for this study. All previously published data are appropriately cited. Code archive DataS1.zip, containing novel R scripts supporting all results, and metadata file MetadataS1.docx, are available at figshare, DOI 10.6084/m9.figshare.16695256.

---

*Corresponding author. Department of Ecology and Evolutionary Biology, E145 Corson Hall, Cornell University, Ithaca NY 14853-2701. Email: spe2@cornell.edu Phone: 607-257-2036, 607-220-3972

The most troubling sentence in Doak et al. (2021, hereafter, D21)) is this: "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". We fully agree that there should not be a "cult of IPM". Population ecologists use matrix projection models (MPM), integral projection models (IPM), differential equations, stage- and physiologically-structured models, agent-based models, and more. All have a place. If your model is a poor choice for your species, data, and questions, or it has been poorly parameterized, that is a legitimate criticism. "Out of fashion" is not.

However, we are not aware of any papers actually "calling into question the usefulness of the many studies based on matrix models" as D21 assert, claiming that IPMs are universally superior to MPMs, or asserting that a "fundamental re-tooling" is needed. The many recent comparative studies using the COMPADRE/COMADRE matrix model databases (Salguero-Gómez et al., 2016, 2015), including work by some of us (Adler et al., 2014), suggest that few ecologists actually question the usefulness of MPMs. So in the published literature, we see no real threat to continued use of MPMs. How editors and reviewers interpret that literature is, of course, a separate matter.

We further agree with D21 that MPMs and IPMs should not be dichotomized; modeling any structured population should involve a common set of considerations and rules of thumb. Despite advocating a more unified modeling framework, D21 perpetuate the idea that these are distinct, competing approaches in their Figure 1, a modeling workflow that forks at an initial choice between MPM and IPM. As a counterpoint we suggest a workflow emphasizing a single stream of considerations from data collection through model analysis (Fig. 1). Depending on the species, data set, and question, our workflow could produce an IPM, MPM, or another structured population model. We use this diagram to frame two central recommendations that represent our primary disagreements with the guidance given by D21:

1. **Use the best available statistical modeling approaches to estimate state-dependent individual survival, growth, and fecundity, and offspring size distributions.** The most important contrast, we believe, is not MPM vs. IPM but statistical demographic modeling versus discretizing continuous states into a set of contingous discrete classes, and using observed class-transition frequencies as projection matrix entries. We refer to the latter as "binning". Indeed, past statements about advantages of IPMs over MPMs are largely about benefits of statistical modeling over binning for continuous state variables – benefits which are often substantial for the questions ecologists are now asking. We do not dispute that statistical modeling and binning usually generated similar results in D21's analyses. Rather, we argue that D21 conducted their analyses under restrictive conditions that do not represent most current applications of these tools. We emphasize key advantages of statistical modeling in more diverse and realistic settings.
2. **Construct the population model that is implied by the vital rate models, and implement it numerically using accurate methods.** In particular, if you use continuous functions to model smooth responses to a continuous trait – which we argue should be default (Fig. 1) – you should increase the number of size classes until all model outputs of interest stop changing in value. We also disagree with D21's advice for discretizing IPM kernels, because (as we show below) no approach is universally "best".

Our main goals are to elaborate on these recommendations, based on D21's case studies and other evidence, and to provide an alternative vision for data-driven modeling of structured populations. With few exceptions, we agree with D21 on the factual content of their paper; our disagreements center on interpretations of that material and the consequent recommendations for "best practices".

Use the best available statistical approaches to estimate state-dependent survival, growth, and fecundity, and offspring size distributions.

At the heart of any structured population model is the relationship between an individual's state and its demographic fate. Whether the state variable is continuous or discrete, we recommend that the state-fate relationship should be modeled statistically (D21's "CVR", for "continuous vital rate" estimation). This is standard for IPMs, in which regression modeling of state-fate relationships forms the backbone of the projection kernel, but it is also possible for classical MPMs based on discrete state variables, as D21 point out (in fact, two authors of D21 were among the first to advocate for parameterizing MPMs through statistical modeling of state-fate relationships (Morris and Doak, 2002)). The alternative, advocated by D21 for some situations, is to estimate transition rates directly from raw data by binning (D21's "DVR", for "discrete vital rate" estimation).

DVR is not inherently wrong, but statistical modeling has at least two advantages. First, as D21 point out, it provides a principled way to account for multiple sources of variation in the data, ranging from nuisance variables that could bias parameter estimates to covariates or experimental treatments that address the primary questions of the study. Statistical modeling makes it possible to identify relevant sources of demographic variation by applying established model selection criteria to compare and choose among competing descriptions of the system. The risks of parameterization via statistical modeling, according to D21, are that simple functions may miss complexities in the data and complex functions may be hard to fit. But these challenges have broadly accessible solutions. Fitting nonlinear smooth functions is a great strength of R, for example, and mixed effects models can be used to obtain unbiased estimates of the main effects of interest in the face of nuisance variables. Suggesting that complexity acts as a barrier underestimates the statistical savvy of demographers, and under-exploits the modern revolution in statistical computing. Fitted regression models require scrutiny, and strongly non-Gaussian growth patterns require thoughtful choice of a better distribution. But there are standard model selection tools and diagnostics for regression models, and work by D21's authors provide nice examples of how to construct more complex growth models and vet them to ensure that they describe the data well (Peterson et al., 2019; Shriver et al., 2012). While CVR requires some foundations in statistics, we do not view DVR approaches as inherently simpler or less error-prone. In fact, a CVR workflow encourages model builders and model users to focus on evaluating the assumptions and adequacy of the model, rather than on the formal mathematics (see Appendix S1: Section S1). Studies have found chronic errors in the assembly of DVR-based matrix models (Kendall et al., 2019; Stott et al., 2010), and the tedious problem of size class boundaries – how many and where – has troubled our field for decades without resolution (Moloney, 1986; Picard et al., 2010; Ramula and Lehtilä, 2005; Salguero-Gomez and Plotkin, 2010; Vandermeer, 1978).

The second advantage is that statistical models can help fill in gaps where few or no values of a state variable or covariate exist. This can be critical when a matrix is estimated from a small number of observations per class. For example, Ramula et al. (2020) used statistical modeling to parameterize 10-class stochastic matrix models for four populations of the long-lived herb *Astragalus scaphoides*, even though population×year-specific sample sizes were as low as seven. In any size-structured population, some sizes are common and others are very rare, but we need to estimate state-fate relations for all of them. If a gap in observed values is too wide, an entire class might be empty, demanding fewer, wider classes or else risking the construction of a matrix that is reducible and/or non-ergodic because the model omits some real transitions. A reducible matrix can have biologically implausible dead-ends in the life cycle. A non-ergodic matrix has multiple stable state distributions, typically with different long-term growth rates, which pose problems for analysis and interpretation. Stott et al. (2010) showed that *ca.* 25% of published DVR-based matrix models include biologically implausible discontinuities in the life cycle (i.e., are reducible and possibly non-ergodic), a consequence of some life-history transitions that actually occur going undetected in a finite sample.

In their analyses to compare CVR versus DVR for small samples, D21 avoid this problem with fully stratified sub-sampling that ensured proportional representation in every size class regardless of sample size – the best possible case for DVR to match the accuracy of CVR. They justify this approach by stating that ecologists often intentionally seek out individuals of different sizes. This is sometimes true, but field scientists cannot ensure that every size class is well-represented if class boundaries are not defined until the data are in hand, as D21 recommend. At best, they may seek out roughly even size representation, but most natural populations look something like D21's Fig. 6: some sizes will always be hard to find, even with targeted effort. We replicated D21's sample size analyses for their *Polygonum viviparum* (bistort) data, but with realistically small samples based on two recent studies (Fig. 2). To be generous, we used only 10 size classes, the minimum recommended by D21, but stratified sampling based on 5 classes; this simulates the process of seeking even size-representation in the field without knowing in advance what the model will be. We found the DVR method often generated reducible matrices when sample size decreased below 300, many having a dominant eigenvalue $\lambda = 1$, which results from an estimated survival rate of 100% for some isolated portion of the life cycle with no chance of reproduction. In contrast, CVR virtually never generated such matrices even at the smallest sample sizes. *Contra* D21, CVR clearly outperforms DVR at small sample sizes.

These two advantages – accounting for sources of variation and bridging data gaps – are related. With DVR, accounting for relevant covariates (e.g., year and site) is done by cross-classified binning, meaning that classes are defined by the values of two or more state variables. This is equivalent to independently constructing many projection matrices (e.g., one for each year×site combination, using only data for that year and site), which reduces sample size per projection matrix and amplifies the problems of parameter uncertainty and matrix reducibility. A CVR approach to the same data could use mixed models in which year- and site-specific random effects "borrow strength" from the full data set.

To highlight these potential advantages of regression approaches and their relevance for contemporary ecological research, one needs to look no further than the papers that generated D21's data sets. The previously published analyses of three of the five case studies in D21 (Doak and Morris, 2010; Fitzpatrick et al., 2016; Shriver et al., 2012) used statistical modeling (CVR) to test driver variables, account for imperfect detection, and accommodate sources of variation (across time, location, or ancestry) that would make DVR with more than a few size classes problematic (see Appendix S1: Section S2 for details). The other two (García, 2003; Linares et al., 2007) did use DVR, but because of cross-classified states, had to use far fewer size classes than D21 recommend. Thus, none of the analyses in D21 correspond to the deeper, more complex questions that really were asked about the featured populations. Those questions demanded cross-classifications that greatly increase the data requirements for binning with the minimum 10 even-sized classes that D21 advise. The domain that D21's case studies represent has limited relevance to the analyses of most interest today, where binning quickly encounters the "curse of dimensionality". In contrast, CVR can use multivariate regression models whose parameter count grows far more slowly as covariates are added.

These observations and the evidence in Fig. 2 lead us to conclude that the equivalence of DVR and CVR emphasized by D21 is a special case that applies only when it is possible to ensure that the data set will include good representation for all size classes (before knowing what those will be): either very large sample size, or a simple model with no cross-classification by multiple state variables and no environmental covariates. The case studies that D21 use to compare DVR with statistical modeling all satisfy these conditions, but most modern demographic analyses and questions do not.

Construct the population model that is implied by the vital rate models, and implement it numerically using accurate methods.

Demographic data analyses aim to identify and quantify the factors governing the fates of individuals, and how they vary over space and time, leading to statistical models that summarize our understanding of state-fate relationships. A structured population model should embody that understanding. In particular, the mathematical model should align with the conceptual model of the population. If all state variables are continuous, aim to build an IPM; if they are all discrete, use an MPM. If they are a mixture, incorporate discrete stages and continuous state variables into a hybrid model. This advice is not controversial. Indeed, D21 adopt a continuous conceptual model whenever they use CVR, even if they call the result a "matrix model" (in fact, they call those models "IPMs" in their supplementary R scripts.)

When only continuous size has been measured, the conceptual model is a simple IPM:

$$n(z',t+1) = \int K(z',z)n(z,t)dz = \int \left[P(z',z)+F(z',z)\right]n(z,t)dz \qquad (1)$$

where P and F are the survival/growth and fecundity components of the kernel K, z is the size measure, and $n(z,t)$ is the size-distribution function for the population at time t. The difference between MPMs and IPMs is that, in the latter, the trait remains continuous in the underlying conceptual model. What D21 say about equivalence of MPMs and IPMs (D21, p. 6) fails to distinguish the model itself from the numerical solution methods: an integral that has been evaluated numerically is still an integral. Midpoint rule (which D21 call the "mesh point method") and what D21 call the "CDF difference" method (see Appendix S1: Section S3)) both use evenly spaced kernel values, and allow us to think of the IPM as a matrix model with narrow size classes and re-use R or MATLAB code for matrix model calculations.

In their Abstract D21 call it a "surprising result" that "the most commonly used method to discretize growth rates"[midpoint rule]... "can introduce substantial error into model outputs" (their Fig. 7b), calling into question the entire IPM literature. But as they later acknowledge (D21 p. 16), it has long been recognized that midpoint rule becomes inaccurate with too few size classes (Zuidema et al., 2010, Fig. 2). Solutions are available (Ellner et al., 2016, Ch. 6). In many cases the solution is simple: use more size classes (Fig. 3B) until the quantity of interest stops changing. At minimum, the number of classes should be large enough that the column-sums of the growth kernel sum to the size-specific survival probability. Other cases – extremely "spiky" kernels (i.e., very small σ_g), or cross-classification by multiple continuous attributes – may require more efficient methods (reviewed in Appendix S1: Section S3 with some recommendations).

D21 advise that CDF difference generally out-performs midpoint rule, but that is not always true. CDF difference is designed to be very accurate for projecting how many individuals fall in each size class next year (conditional on survival). But that is only one of many things we do with structured population models.

We also use discretized IPM kernels to compute generation time, selection gradients, reproductive skew (the variance among individuals in lifetime reproductive success), and so on. Perhaps surprisingly, CDF difference may not be best for every purpose. For example, a key calculation for analyzing among-individual variation in lifetime reproductive success (Snyder et al., 2021) is the variance of expected future reproduction as a function of size at time $t+1$, conditional on size at t. For the growth distribution in D21 Fig. 7, suppose for illustration that expected total future reproduction is proportional to $\exp(z/50)$, where z is current size (the actual formula is more complicated and must be calculated numerically). Midpoint rule is very accurate with 200 size classes, but 300 classes is nowhere near enough for CDF difference (Fig. 3C). These results also illustrate the crucial point that different calculations require different numbers of classes: midpoint rule with 150 classes is fine for total survival, but not for variance in expected future reproduction. Appendix S1: Fig.S2 illustrates in a real, data-driven model that the relationship between class number and accuracy varies enormously depending on the model and on what quantity is being computed.

So the essential points here are not about midpoint rule vs. CDF difference. They are that first, no discretization method is best for everything, and second, there is no generally reliable "rule of thumb" that 10, 30, 200, or any other number of size categories is usually sufficient for numerical accuracy. The Royal Society motto "Nullius in verba" (take nobody's word for it) is a better rule of thumb. Practitioners should be aware of the options and their relative merits. We like midpoint rule because it is simple, robust, and often does the trick; when it fails we try something more sophisticated. But whatever method or methods you use, it is essential to verify that answers don't change meaningfully when more size classes are used. This is straightforward with a CVR-parameterized model – varying class number should be a trivial code edit. With DVR it is problematic, because sampling variability will cause answers to continually change, rather than converge, as the number of size classes is increased. Deciding when "enough is enough" remains seat-of-the-pants.

Conclusions

We fully agree with D21 that differences between MPM and IPM have sometimes been overstated. We have argued (and the authors of D21 may well agree) that the more important dichotomy is traditional binning versus statistical modeling for model parameterization – for continuous traits, what D21 call DVR vs. CVR. D21 have shown, and we do not dispute, that a binning-based matrix model can be as effective as an IPM for size-structured populations if it becomes a whole lot more like an IPM by using narrow, evenly-spaced size classes instead of the historically predominant wide, uneven classes. However, actually constructing such a model from empirical data is possible only in restricted circumstances where it is feasible to ensure that all of those many classes are well-represented in the data set, as is true in D21's simulations. This is hard to achieve even in simple models with current size as the only predictor of future size, and much harder with continuous covariates or cross-classification. It was conceptually useful for D21 to distinguish three different modeling decisions (DVR vs. CVR, number of classes, even vs. uneven classes), but by focusing on the "main effect" of each decision, D21 under-emphasized an important interaction among them: even in ideal circumstances for DVR, created by pooling data across important heterogeneities, CVR with many even classes is sometimes matched but never bettered. Moreover, CVR is often essential for the multivariate analyses that ecologists now want to do, such as relating demographic rates to habitat or environment variables to allow projections about no-analog future conditions. For continuous traits, CVR also solves the longstanding problem of choosing the number of size classes and their boundaries – just increase the number of classes until answers stop changing. Then you can call it what you want, or whatever your audience is most comfortable with, but what you've built is an IPM.

Acknowledgements All authors discussed all aspects of the research and contributed to writing and revising the paper. Research supported by NSF grants DEB-1933497 (SPE, GH), DEB-1933561 (PBA), DEB-1754468 (TEXM). We thank Dan Doak for providing unpublished R scripts and Yngvild Vindenes, the authors of D21, and an anonymous reviewer for helpful comments on the original manuscript.

Literature Cited

Adler, P. B., R. Salguero-Gómez, A. Compagnoni, et al. 2014. Functional traits explain variation in plant life history strategies. Proceedings of the National Academy of Sciences, **111**:740–745.

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., E. Waddle, R. E. Langendorf, et al. 2021. A critical comparison of integral projection and matrix projection models for demographic analysis. Ecological Monographs, **91**:e01447.

Ellner, S., D. 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., J. C. Gerberich, L. M. Angeloni, et al. 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.

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

Kendall, B. E., M. Fujiwara, J. Diaz-Lopez, et al. 2019. Persistent problems in the construction of matrix population models. Ecological Modelling, **406**:33–43.

Linares, C., D. F. 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.

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

Morris, W. F. and D. F. Doak. 2002. Quantitative Conservation Biology. Sinauer, Sunderland, MA.

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., D. Ouédraogo, and A. Bar-Hen. 2010. Choosing classes for size projection matrix models. Ecological Modelling, **221**:2270–2279.

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.

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

Salguero-Gómez, R., O. R. Jones, C. R. Archer, et al. 2016. COMADRE: a global data base of animal demography. Journal of Animal Ecology, **85**:371–384.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, et al. 2015. The COMPADRE plant matrix database: an open online repository for plant demography. Journal of Ecology, **103**:202–218.

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

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.

Snyder, R. E., S. P. Ellner, and G. Hooker. 2021. Time and chance: Using age partitioning to understand how luck drives variation in reproductive success. The American Naturalist, **197**:E110–E128.

Stott, I., S. Townley, D. Carslake, and D. J. Hodgson. 2010. On reducibility and ergodicity of population projection matrix models. Methods in Ecology and Evolution, **1**:242–252.

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

Zuidema, P. A., E. Jongejans, P. D. Chien, et al. 2010. Integral projection models for trees: a new parameterization method and a validation of model output. Journal of Ecology, **98**:345–355.

Figure legends

Figure 1. Recommended workflow for construction and analysis of structured population models. Dashed arrows represent iterative feedbacks: fitting, vetting, and re-fitting vital rate models until they describe the data adequately and increasing class number in model construction until outputs stop changing.

Figure 2. Proportion of matrices that **A)** are irreducible and **B)** have population growth rate $\lambda = 1$ exactly, constructed from stratified subsamples of the bistort (*Polygonum viviparum*) data of D21. Vital rates were estimated using either binning (DVR) or continuous regression-based parameterization (CVR). We used 10 classes to construct the models and subsampled using 5 classes, which guaranteed that all samples included data from all 5 classes. The shaded area indicates the distribution of sample sizes used by Ramula et al. (2020) and Louthan et al. (2018) to estimate projection matrices (81 total). This demonstrates that many field studies are in the danger zone where DVR-parameterized models are at risk of being reducible, with dominant eigenvalue describing an isolated part of the life cycle with 100% survival and no reproduction. Figure made by scripts DataS1: `bistorts STRAT-RARIFY-bins.R` and DataS1: `Stott MEE.R` using R version 4.0.2.

Figure 3 Comparisons of numerical integration accuracy. **A)** and **B)** correspond to D21 Fig. 7, a Gaussian growth distribution with size range [0,100] where size next year is Gaussian with $\sigma = 0.5$ and a sequence of means ranging from 49 to 51. **A)** The sum of growth probabilities (with true value 1) calculated by midpoint rule, exactly as in D21 Fig. 7b, showing large errors when size classes are too wide. **B)** As in A) but with more size classes, showing that the error becomes trivially small. **C)** Calculation by midpoint rule (dashed black) and CDF difference (solid red) of the variance of expected future reproduction (with expected future reproduction assumed to have functional form $25\exp(z'/50)$) as a function of size z' next year, for initial sizes 48 and 52. Dashed black horizontal lines are the values using 1000 size classes with midpoint rule. Figure made by script DataS1: `Convergence_comparison.R` using R version 4.1.1.

Figures

Figure 1:

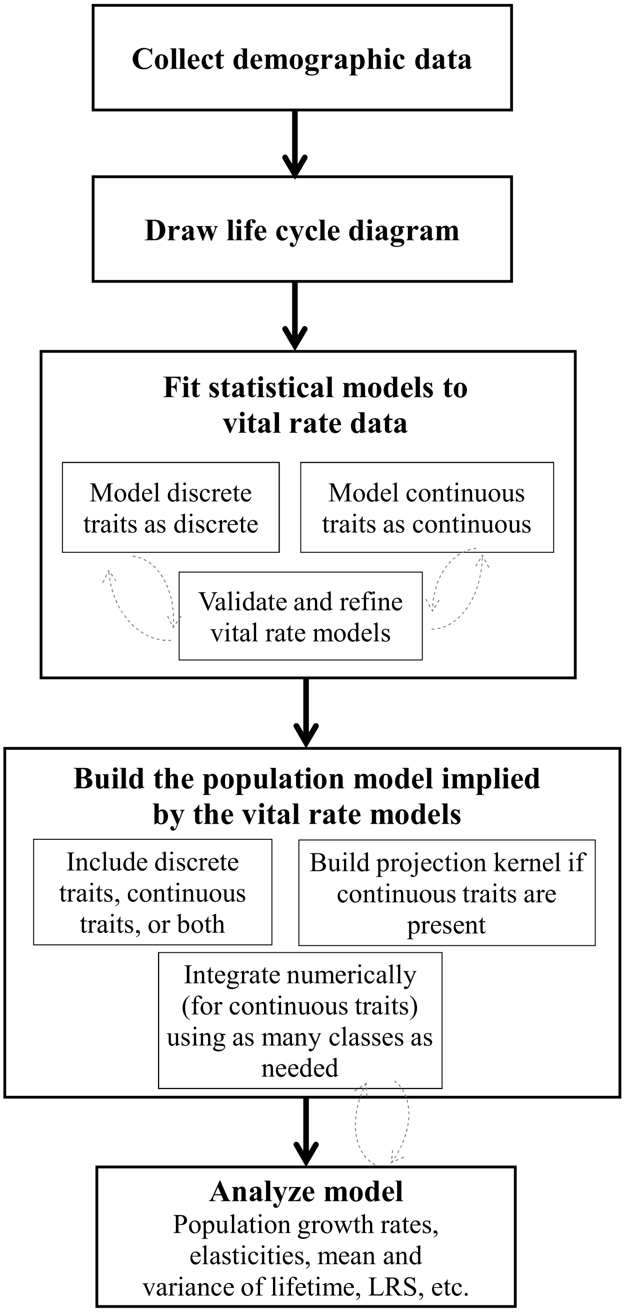


Figure 2:

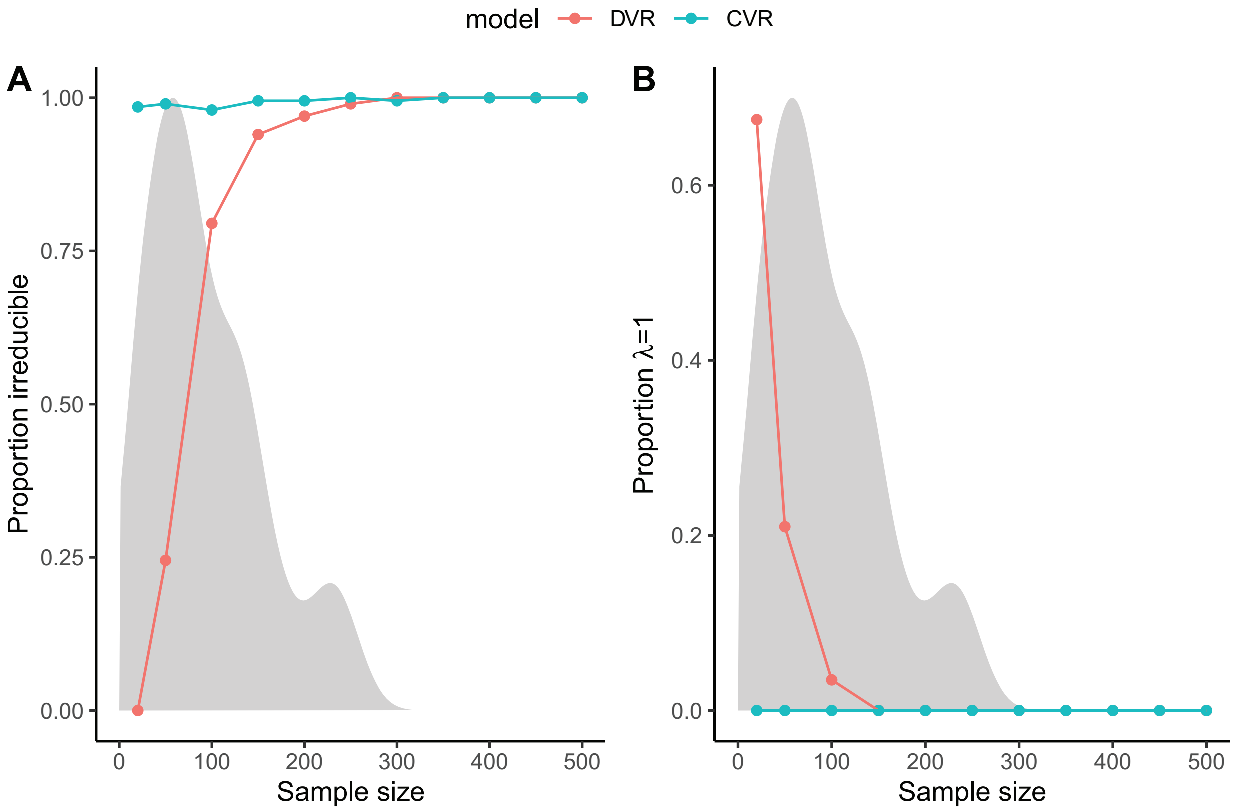
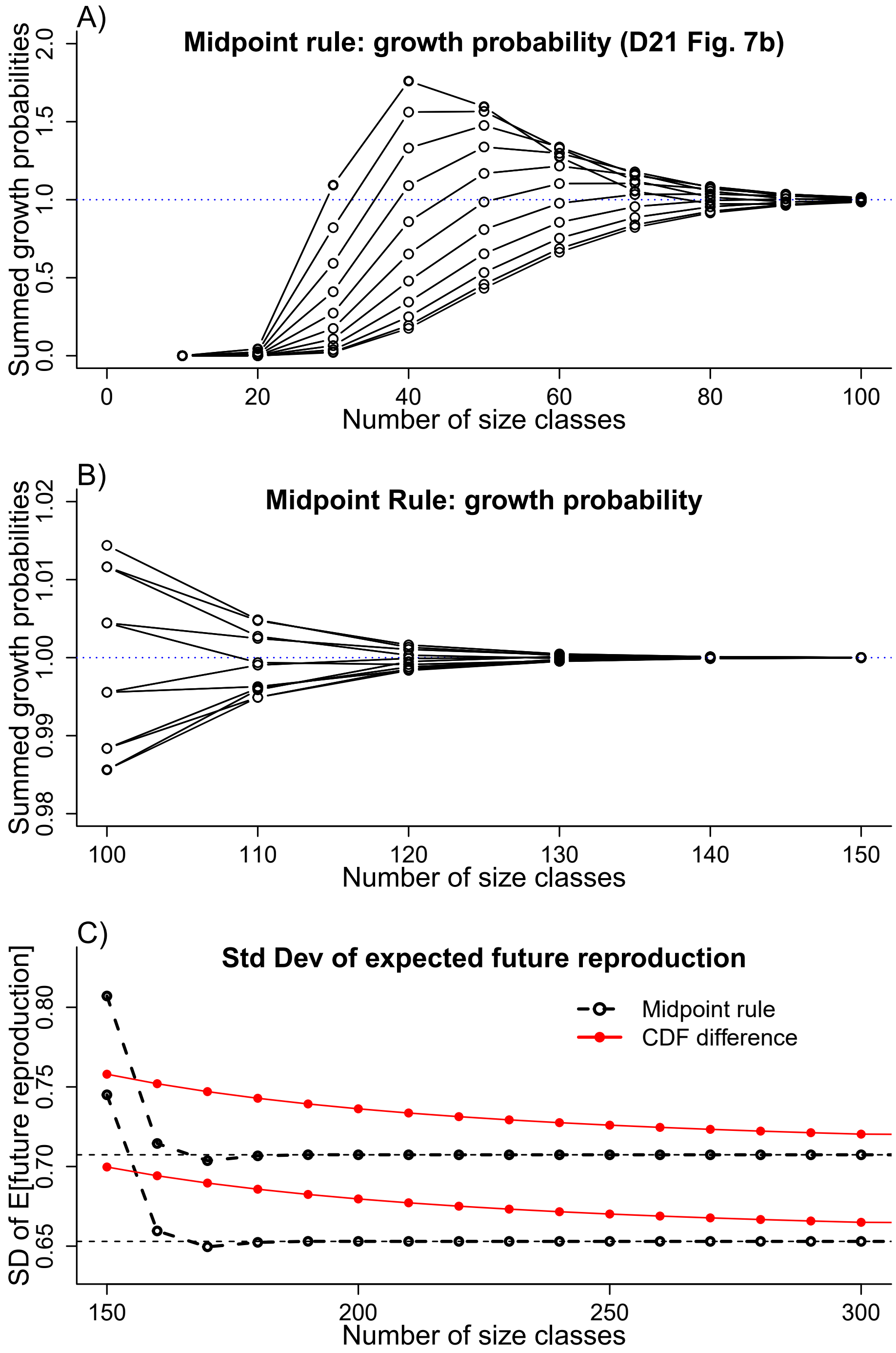

model
DVR
CVR
A
Proportion irreducible
1.00
0.75
0.50
0.25
0.00
0
100
200
300
400
500
Sample size
B
Proportion λ=1
0.6
0.4
0.2
0.0
0
100
200
300
400
500
Sample size

Figure 3:


A)
Midpoint rule: growth probability (D21 Fig. 7b)
Summed growth probabilities
0.0
0.5
1.0
1.5
2.0
0
20
40
60
80
100
Number of size classes
B)
Midpoint Rule: growth probability
Summed growth probabilities
0.98
0.99
1.00
1.01
1.02
100
110
120
130
140
150
Number of size classes
C)
Std Dev of expected future reproduction
SD of E[future reproduction]
0.65
0.70
0.75
0.80
Midpoint rule
CDF difference
150
200
250
300
Number of size classes