Response to reviews --- ECOMOD-18-618

Reviewer comments in black

Author responses in blue

**Reviewer #1:** This manuscript ("Persistent problems in the construction of matrix population models") presents an overview of a class of common errors in constructing matrix population models (MPMs) from data.  Specifically, the problem is that users of such models often do not include a full year's worth of demography in "annual" transition rates (or, similarly, do not include a full week of demography in a weekly model, etc).  I think the paper has the potential to make a valuable contribution to the field, though it is still rough in many ways (as outlined below).

The paper has two clear strengths.  First, I appreciated the clarity with which these common errors were explained, and the reference to learning theory and why such errors may be so pernicious.  Second, the article reviews published matrix models for animal populations, and shows that such errors are common (present in a bit more than half of the published models).

These were two major goals of the paper, and we are pleased that the reviewer finds them clear and valuable.

The weaknesses of the paper are in the exposition of the case studies, and, especially, in the discussion of how users of models could improve their use of them.

See our responses to the detailed comments below.

The problems with the case studies are twofold:  First, the main text is a long list of results with reference to figures in the appendix.  I find this difficult to read and fear that people who don't like models will zone out and completely miss this section.  I suggest that the authors consider one or two points they want to make about the contrasting models and present them as figures.  I would be tempted to suggest comparing the growth rates of the three populations using an LTRE-like analysis, i.e., as you add each correction (in turn) to the original model, or each error (in turn) to the correct model, how much does the growth rate change?  But, the main point is to make sure there are one or two clear take-home messages for readers who are uncomfortable with models.

This is an excellent suggestion. We have highlighted the impacts of matrix construction errors on estimates of lambda and on the elasticity analysis of lambda, completely reorganizing the section (now section 4) and adding two new figures (5 and 6) to the main text that show the interacting effects of the three sources of error (while we don’t explicitly call it an LTRE, it is very LTRE-like -- thanks for the suggestion!). We then retain a much briefer description of impact on the other analysis endpoints, referring to figures in the appendix.

Second, the text nearly entirely focuses on what the erroneous models get wrong.  There should be equal attention paid to what the models get right. For example, the (erroneous) published models correctly identify whether each population is growing or declining.  They also identify gross differences in sensitivities and elasticities of different life stages. Thus, they are useful tools for their most basic management applications, even when done incorrectly.  This robustness to human error is an important message for users of the models - it is better to get the models right, of course, but, even in the presence of these common errors, the models are picking up basic aspects of the life cycle that might not be apparent without a model.   I suggest adding this point to the results comparing models, and to the discussion in the form of advice to empiricists (see below).

We have added a paragraph (ll. 616-31) to the Discussion pointing out that, in the case studies, the “sign” of the growth rate, the rank order of stable stage and reproductive distributions, and (sometimes) the rank order of elasticity contributions were not affected by the matrix construction errors. However, we are less optimistic than the reviewer, in large part because our case studies are too few to fully span life history space. Thus, we hedge these observations on “what the models get right”:

*Nevertheless, the case studies we analyzed suggest that many analyses conducted with flawed MPMs may have broad validity. For example, in none of the case studies did the errors introduce misclassifications of population decline versus increase. Furthermore, errors in $\lambda\_1$ are generally modest, except when an inappropriate maturation model is used (Fig. 5). Finally, the rank orders in stage-specific values in the stable stage  and reproductive value distributions were unaffected by the matrix construction errors. However, we have only a limited sample of case studies, and for some outcomes even this small sample gives inconsistent results for the direction or magnitude of biases introduced by the MPM construction errors. For example, the rank order of the elasticities of lambda to the vital rates might (alligator) or might not (lionfish) be robust to model errors. While a more complete understanding of these biases might be derived by analyzing more models, spanning a range of life histories, ecological understanding and management decisions would be better served by constructing the models correctly in the first place.*

The problem with the Discussion is that the advice to users of models ("prebreeding  census models are less prone to errors") is not useful, seems slightly condescending, and is possibly misleading.  In my experience, one's ability to structure a model in different ways is far more flexible than one's ability to collect data in different ways.  I imagine a scenario in which a user of matrix models who doesn't understand their construction learns that "prebreeding models are better" then tries to twist their available data from a post-breeding census study to fit a prebreeding models.  Perhaps the authors of this paper have different experiences that suggest greater flexibility in data collection. But, to me this seems like strange advice.

Our “advice to users” was insufficiently nuanced, and we appreciate both reviewers for challenging us on this section. But before describing our revision, we want to challenge the assertion that pre-breeding and post-breeding census models require different data. For a population in which *either* formulation is appropriate (i.e., the breeding season is very short, such that adult mortality within the breeding season can be ignored), then **the data requirements are identical** for the two types of model. That is, one needs an annual survival for each age or stage class, and an average birth rate at each age or stage. Survival measured from the beginning of one breeding to the beginning of the next will be the same as survival measured from the end of one breeding season to the end of the next. The question of mortality within the breeding season is an important one (though beyond the scope of this paper), but the standard models, and essentially all empirical models, assume it to be negligible. If that assumption is valid, then the survival and birth rate parameters are identical between the two representations, and all appear in both (for example, compare Figs. 1b and 1d in our manuscript).

So our basic recommendations remain unchanged, but we have provided more context and explanation. We use the results of our analysis to motivate the recommendation to construct prebreeding models and provide a hypothesis for this outcome. We follow this with an explanation, similar in content to the above, describing why the same data can be used for both formulations. This motivating paragraph reads (ll. 692-710):

*First, we note that none of the prebreeding census MPMs that we examined had errors in the fertility coefficients. While we cannot assign causality to this outcome based simply on the data, prebreeding models simply do not require dealing with the troublesome reproductive transition by maturing juveniles, and it may be that the fact that newborn survival is "left over" after constructing the survival transitions encourages a closer look at the fertility coefficient. It is also important to recognize that, if one has information on age- or stage-specific birth rates and annual survival, then that information can be used to create \*either\* a prebreeding or a postbreeding census model (e.g., compare Figs. 1b and 1d, noticing that the same parameters occur in both). In particular, unless there is large age-dependent variation in mortality through the breeding season, it doesn't matter whether the survival is measured from the end of one breeding season to the end of the next, or from the beginning of one breeding season to the beginning of the next. The key is that the survival be measured over a full timestep (e.g., year). Furthermore, the prebreeding and postbreeding censuses are different views into the \*same model\*. In particular, the estimate of $\lambda\_1$ and the elasticities of $\lambda\_1$ to underlying demographic parameters will be identical between the two representations.*

The next paragraph retains much of the content from the previous version of the manuscript, except that we have emphasized the purely mechanical nature of moving between prebreeding and postbreeding census representations, ending the paragraph with (ll. 720-5):

*In addition, for the simple sorts of models represented by Figs. 1 and 2 (e.g., without clonal reproduction or multiple genders), the conversion of a matrix from a prebreeding representation to a postbreeding representation is a purely mechanical process that can be automated in software. The pre\_to\_post function in the mpmtools R package (Kendall, 2019) provides an example of this.*

We have also expanded the recommendations for stage structured models, providing motivation/context and taking advantage of subsequent advances in the mpmtools package. Prior to recommending the “unrolled” model, we state (ll. 726-35):

*Second, we found that most models in which a conceptually age-structured population had been collapsed into a small number of stages, each with a well-defined mean stage duration, were constructed in ways that would fail to reproduce the asymptotic growth rate---indeed, in our case studies this error had the largest impact on $\lambda\_1$ estimates. We interpret this as being because the best stage-based approximation---the AAS model---is both conceptually and computationally challenging to produce. For simple life histories, we recommend using the \*\*mpmtools\*\* R package (Kendall 2019) to automatically construct an AAS model from information on stage-specific demography and duration.*

Here are the lessons I would give to users of matrix models, based on this study: (1) The appropriate life cycle for a matrix model is often (perhaps usually) not the cartoon life cycle that you first want to draw for a species.  Each transition has to represent a whole year. Think through what happens to individuals in each stage class over the course of the year - perhaps even drawing out submodels with sub-annual transitions for each stage class. Once you have figured out what happens to individuals in each stage class over the course of a year, then you can create the arrows and appropriate classes for the annual transition model.  [When I teach courses, I often ask students (or empirical ecologists) to think of this as a really hard 4th or 5th grade word problem. It makes them laugh and - hopefully - makes the notion of thinking through the steps less intimidating.]

We agree that effective teaching is key. However, the learning process described by the reviewer, which in our experience unfolds over multiple hours of highly interactive teaching, cannot be replicated in a few discussion paragraphs. The existing texts do not make this learning available to someone trying to do self-study; indeed, we suspect that many population ecology courses, when taught by instructors whose expertise is in other parts of the discipline, rely on existing, inadequate, textbooks to guide their pedagogy. We address this with a new paragraph (ll. 673-86):

*How can the prevalence of these errors in MPM construction be reduced in the future? The standard recommendation from experts (including both reviewers of this paper) is that biologists should think carefully about what happens to individuals over the course of a timestep, think carefully about how the data relate to these events, and assemble everything in such a way that each transition spans a timestep. In essence, this advice boils down to "think like a modeller." In our experience, non-modellers require a lot of interactive tutorial examples to really learn this process. Current textbooks do not provide this; the best published examples we have seen are in articles describing modeling approaches for migratory species (e.g., Fig. 28.2 of Runge and Marra, 2005). It is impossible to know how many instructors provide effective classroom teaching on this topic, but our analysis of the COMADRE database indicates that a substantial number of biologists have not adequately internalized this message.*

We do call for experts to develop more effective training materials. Our advice to non-experts here is meant to provide some practical workarounds in the absence of such training material.

The practice of developing submodels is, we think, a key step--but one that is generally absent from existing texts. Essentially, this involves building seasonal matrix models, which is generally treated as an advanced topic that is reserved for models that explicitly want to track within-year dynamics. Tutorials and texts that build out this approach would be very useful. We have added a paragraph to that effect (ll. 787-802):

*One pedagogical approach that may be helpful is to build "seasonal" submodels, and use matrix multiplication to generate the integration over the full timestep (rather than requiring the integration to be done in the brain of the model builder). Williams et al. (2002) do this (for age-structured models only; section 8.4.2), creating "survivorship" and "reproduction" matrices and showing how they can be combined to produce both prebreeding and postbreeding census models.  Periodic matrix models are generally presented as advanced topics (the notation in Caswell's 2001 treatment of the topic is intimidating to non-modelers), and are typically applied to either annual species or species that breed continuously (at seasonally varying rates) throughout the year. Nevertheless, a more accessible presentation of periodic matrix models, showing how they can be used to integrate demographic processes that happen at different points in the annual cycle, might help alleviate many of the problems identified here. Such an approach to MPM construction might also encourage the development of models that more fully incorporate the richness of within-year life history processes.*

(2) The models are broadly robust to error - do not be afraid to use them after reading this paper.

See our detailed response on model robustness above. We have moderated the cautionary sentence in the final paragraph of the discussion (ll. 839-43) to explicitly acknowledge that existing analyses *may* be ok, but we remain cautious:

*While some analysis endpoints seem to be broadly robust to matrix construction errors (especially errors in the fertility coefficient), there may be some cases in which the scientific conclusions or management recommendations supported by those models could be flawed.*

Having (as reviewers typically do) spent more time on the negatives than the positives of this paper, I want to mention that I really appreciated reading the advice to "experts" teaching or explaining matrix models.  I had not thought about it in this way before, but I think the authors of this paper are correct that most texts have a reasonable presentation of age-structured models, but a shorter and muddled presentation of stage-structured ones.  However, stage-structured models are much more common. I also really appreciated the point that, in order to correct conceptual errors - e.g., the notion that because eggs are a conspicuous life stage they should be a stage in the matrix models, and the notion that pre-reproductive juveniles should never have a fecundity transition - we (the experts) need to start by first acknowledging the common mistakes and talking about them.  I think this paepr will go a long way towards starting that conversation.

We feel that this is an important message of our paper, and we are pleased that the reviewer finds the manuscript effective in that regard.

A few more minor points:

1. Consider reordering the sections to put the review first and case studies second (I think it will flow better, and emphasize that these case studies are typical - not exceptions)

We see arguments in favor of both orderings--our thinking was that the demonstration that the errors have consequences would motivate the need for the review. But we don’t feel strongly about that, so we have done the suggested reordering. In accordance with this, we have reordered the penultimate sentence of the Introduction (which outlined the rest of the paper)

2. I found it confusing to have the two alligator populations on one figure in the appendix.  It made it look superficially like model construction (on the x-axis) had a huge effect on model outputs.  On a more careful inspection, it becomes obvious that the different "models" are actually different populations and the models are correctly identifying differences between these populations (L1-6 consistently differs from L7-12).  But, it would be good for the figure to send the correct message on first inspection

We have separated the two populations into separate panels in each appendix figure (Figs. A.9-A.14), as well as in the new fig. 5 of the main text. We have also given separate abbreviations to the models of the two alligator populations (AN and AS).

3. This may be an Elsevier problem, but I did not notice Appendix B online

We don’t know what the reviewer view to the Elsevier site looks like, but we did upload Appendix B, and, at least from the author view, the combined pdf of the manuscript generated by Elsevier has a link to the appendix on the final page.

4. It seems like "threshold concept" is first used on line 54, then on 624, but defined on line 628.  Can the term be defined when first used?

We have added the definition at the first mention in the Introduction (l. 55), and then reminded the reader of the definition where threshold concepts are revisited in the Discussion (l. 651).

5. The paragraph about plant populations is underwhelming and does not add anything to the paper.  It should be deleted to make more room for discussing other points, e.g., strengths as well as weaknesses of models as currently (and erroneously) used.

We agree that the existing paragraph was problematic, as it essentially tried to summarize a hypothetical manuscript with similar complexity to the current one. However, we do not want to remove all mention of plants, as the question “What about plant MPMs?” is an obvious one (indeed, one made by several readers of pre-submission versions of the ms). Thus, we have replaced the existing paragraph with a much shorter one that reads (ll. 827-33):

*While our analysis has focused on animal MPMs, we note that plant MPMs may be subject to similar errors as those described here. In addition, the construction of size-structured MPMs faces challenges that are qualitatively similar to those for stage-structured MPMs, while differing in detail; and ambiguities around the treatment of seeds and seed banks create the potential for additional errors in the fertility coefficient. Thus, a systematic analysis of published plant MPMs may be warranted.*

**Reviewer #2:** This well-written paper convincingly addresses three important errors that are often made in building matrix projection models from vital rate parameter estimates.  Two of the errors concern the reproductive elements in the matrix top row: failing to include a survival term along with the fecundity term, and failing to allow reproduction for classes that reproduce at the end of the (post-birth pulse) time step.  The 3rd error concerns incorrectly partitioning maturation from one stage to the next (survival and remaining in stage vs survival and advancing).

Not only do the authors clearly explain the problem, they also estimate its prevalence and describe the consequences.  Using 2 real-world published life histories (lionfish and alligators), they show in what direction and how much these errors could affect estimates of asymptotic lambda, age distribution, generation time, sensitivities/elasticities, and so on.  Clearly, these are problems that basic and applied ecologists need to get better at avoiding.

We agree!

    I enjoyed this paper very much and really have just one disagreement to raise. On page 23 line 649 the authors recommend that the non-intuitive reproduction problem inherent to post-breeding formulations could be fixed by instead constructing the model  in a prebreeding framework. This is a bad idea because ultimately the choice of pre vs post breeding matrix formulation should be driven by how the data were collected to derive the estimates of reproduction and survival. If the vital rates were in fact collected just after the reproductive pulse (post-breeding), then the estimates of reproductive output (eg 'fecundity') and survival of newborns would be invalid in a pre-breeding framework.  Even if they could be finessed in a work-around to be equivalent, this solution would encourage sloppy connections between how field data are collected and how they are parameterized in a matrix projection model. Better to force students and practitioners to think carefully about how the data were collected and how to incorporate those data and the organism's life history in a projection model.

Please see our extended responses to a similar criticism by reviewer 1, which describes our revisions. The reviewer’s recommendation is essentially to provide better training material, which is something we do recommend in our advice to experts. Our goal in this section of the manuscript is to provide an accessible workaround that can be used in the absence of such training materials.

Regarding the validity of the recommendation to construct a prebreeding census model: As long as the definition of “newborn” is treated consistently between birth and survival (i.e., the birth rate is “number of newborns in year t produced by an adult alive in year t” and newborn survival is “survival from newborn to age 1”), then the birth and survival parameters are interchangeable between the the prebreeding and postbreeding census models. There are, of course, differences in the constitution of the stage vectors at the end of the model timestep; but almost all published MPMs are focused on estimating lambda\_1 and its sensitivities, which will be identical between the two ways of representing the population.

Note that if the definition of “newborn” is not consistent in the way described above, then the estimated parameters will not be appropriate to *either* model!

A few small edits:

Line 110: "the are"

Corrected this to “they are”

Line 177: The Mills text came out in 2nd edition in 2013.  Figure 6.3 in the 2013 edition is equivalent to 7.3 in the 2007 first edition.

We have updated this, with a cross reference to the figure in the first edition.

Line 405: "The the"

We have deleted the repeated word