

Recommenders:

We have received two reviews on the manuscript, and both are positive. The manuscript requires a relatively minor revision that will not require re-review – the co-recommenders can make the call. We consequently expect to accept the revised manuscript.

The manuscript simulates data typical of those collected in a demographic field study of interacting species and then uses integrated population modelling (IPM) to recover the parameters used to simulate the data. The work is sound, and the manuscript is well written. The reviewers have identified some minor edits that will aid clarity.

One reviewer raises two slightly more major concerns. First, the reliance on a previous publication. This manuscript builds on that work, so I am not overly concerned by this. However, a few edits to make this manuscript a little more stand alone would not go amiss.

The second issue is one would hope that IPMs would recover parameters from simulations when the same model structure is used to simulate and estimate the process model in an IPM. These tests are consequently not particularly challenging. I do think this needs more acknowledgement in the discussion. In reality, we never know the generating process when we collect data, and we have to make decisions on what the process and error models will look like in an IPM. Get one of the wrong, and results can be unreliable. I am not going to request more simulations here, but in future work it might be interesting to select a broader range of models that differ from the simulation for the process model of the IPM. Perhaps stress this point more.

Authors' response:

We thank the recommenders and the reviewers for their constructive comments and suggestions.

First of all, following a comment from a reviewer, we realised that the way our capture-recapture model was specified was not allowing for individuals to be recaptured more than once. That is, we used a capture-removal model whereby individuals are marked as juveniles and later recaptured and removed from the study population (e.g., fished, hunted or released elsewhere). Technically, the m-array datasets for the adults were missing. We have now corrected this and reran all analyses. Results remained very similar, except that (logically) precision in survival parameter estimates is now better (since there is more data with adult recaptures), and the vast majority of the models now converge (from 92% to 100% depending on the scenario), including when using the smaller samples sizes considered in Barraquand & Gimenez 2019, and when including random temporal variation in the vital rates. We have revised the text where needed to fit with these changes (and of course updated all figures).

Regarding the first issue, we have now substantially reduced the number of references to the previous study where not needed (e.g., lines 79, 110, 150). Regarding the second issue, while we completely agree that using data-generating models that are different from the models being fitted is a logical next step, simulating using the statistical being fitted is a necessary first step as it is required to evaluate all statistical properties of the model, which includes (i) the identifiability of model parameters (and assess the amount and type of data needed for practical identifiability), (ii) the coverage of parameter estimates (by definition, coverage will be non-nominal if the two models are different), and (iii) estimator bias, as bias in parameter estimates can occur even in cases where both models are identical (as in e.g. estimators for population variance, which are typically corrected to avoid bias).

We now better highlight these useful features of our approach in the discussion lines 294-297, and encourage future work to explore the robustness of predator-prey models to a range of credible model misspecifications (citing work from Plard et al 2021 doing this for one-species IPMs).

Reviewer 07 Apr 2023 10:16

1-observation error

The authors use a lognormal distribution for the observation error. This generates non-integer observation values for population counts. This seems to me unrealistic. I guess that rounding to the nearest integer might probably not change much the results, or is there a reason for not doing this? (a comment on this in the method section would be sufficient to me)

Authors response:

The use of Gaussian or log-Normal errors to model count data is a customary choice in IPMs and state-space models more generally (Besbeas *et al.*, 2002; Dennis *et al.*, 2006; Auger-Méthé *et al.*, 2021). Poisson models are an alternative but often do not model well the observation error (under or overdispersion). Rounding to the nearest integer is problematic because it is a non-probabilistic process: this makes impossible to fit the model that is being simulated. A Poisson Log-Normal model (Poisson sampling overlaid on top of a Log-Normal distribution, which allows for overdispersion, and is very similar to the Negative Binomial) would be feasible but a bit of an overkill here. The influence of count data discreteness typically arises for very low counts (Warton, 2018) around an average of 1 (i.e., with many zeroes), while counts considered here are much larger. Thus although it is not perfectly “realistic”, the fact that some counts of individuals are 37.2 rather than 37 or 108.6 rather than 109 does not affect model behaviour in a substantial way.

As suggested, we now comment on this in the Methods lines 105-107 “Other choices of observation model

are possible but this one is standard for abundance values that are not too small (Besbeas *et al.*, 2002; Dennis *et al.*, 2006).”.

2-capture-mark-recapture data

I am not sure to get the exact meaning of the $m_{t,j}$ values presented in paragraph 2.1 (for $t < j < T$). Is it the number of juveniles marked at time t that have been resighted at time j for the first time (ie, not resighted before j)? In this case, maybe rewrite the equation for $\theta_{t,j}$ by putting $(1-p)^{j-t}p$ before the product Π , otherwise it is unclear whether this term is within this product or not. But then, why not recording the complete history of re-sighting? For instance a juvenile marked at time t , re-sighted at time $t+2$ and time $t+5$ would appear in $m_{t,t+2}$, but the fact that it has been re-sighted at $t+5$ would not be used in the inference, while it is informative on the survival probability and on the probability of resighting. Or am I misunderstanding something, here?

Authors response:

We are very thankful to the reviewer for their comments, as they made us realise that we did not model the capture-recapture data in the way that we intended to.

Regarding their first comments, the reviewer is right that we should have put $(1-p)^{j-t}p$ before the product Π in the absence of brackets. To keep the chronological order of events within the product of probabilities, however, we preferred using parentheses so that

$$\theta_{t,j}^J = \phi_t^J \left(\prod_{k=t+1}^j \phi_k^A \right) (1-p)^{j-t}p.$$

Regarding the second point, the reviewer is also correct and we indeed realised that the way our capture-recapture model was specified was not allowing for individuals to be recaptured more than once. That is, we used a capture-removal model whereby individuals are marked as juveniles, later recaptured, and then removed from the study population (e.g., fished, hunted or released elsewhere). Technically speaking, we forgot to include the m-array datasets for the adults. We have now corrected this error and re-ran all analyses (see our response to the recommenders). Importantly, we now describe the m-array format and its underlying model in much more detail to improve clarity (see lines 110-142).

That said, the error had at least the collateral benefit to also produce results for capture-removal data. We therefore mention lines 253-259 in the discussion that capture-removal data would produce similar results but with less straightforward model convergence. This is also mentioned now in the code repository

(https://github.com/MatthieuPaquet/multi_species/blob/main/removal/README.md).

3-cases without MCMC convergence

I guess that people analyzing empirical datasets and facing convergence issues would try to lengthen the chains to reach convergence. I wonder whether for such datasets the likelihood of obtaining false positives would be roughly the same or whether such problematic cases are also associated with larger probabilities of false positives?

Authors response:

This is a fair question. In the previously submitted manuscript, in cases where convergence was not reached, the patterns of the chains indicated a lack of parameter identifiability (there was no hope to ever reach convergence). In such cases, lengthening the chains would likely still not allow convergence. Since we re-ran our models with added adult recaptures, we no longer face substantial convergence issues with any model specifications (*a minima* 92/100 of the models showed satisfactory convergence and mixing). Therefore, there are now not enough cases without satisfactory convergence to directly compare them with cases that converge. What we can say, however, is that parameter estimates and coverage values were equally satisfactory when including MCMC samples for all 100 models and when excluding the very few model runs that did not show satisfactory convergence and mixing (from 0 to 8 depending on the scenario).

4-model misspecification

This simulation study is performed in the favorable case in which data are simulated with the model used for the inference process (ie, no model misspecification). While this is a sensible way to proceed, I wonder whether the authors could elaborate some thoughts on the likelihood (or not) of increased occurrence of false positives for certain types of model misspecification. I note that they are somehow touching on this issue in their last paragraph when recommending to reiterate their approach to "new systems with different life history parameters and density-dependent structures". But if they could provide readers with some general (untested) predictions or warnings, this could help.

Authors response:

We have added some more words on this topic at the end of the discussion (lines 298-303). We mention that "an important next step will be to evaluate the sensitivity of multi-species IPM estimates to model

mis-specifications (Plard *et al.*, 2021). For example, different functions than the log and logit links chosen here may be used to fit or to simulate intra- and inter-specific density-dependencies. Hence, we encourage future work to try fitting a broader range of plausible models that differ from the model used to simulate the data (or conversely, to simulate from more mechanistic models) in order to assess such sensitivity. ”. Indeed, in addition to changed density-dependent structures (e.g., predators also eating adult prey, which we already mentioned), the sensitivity to the very functions used for density-dependence could be evaluated.

While we completely agree that a logical next step would be to investigate model mis-specifications, and as the reviewer can see, we are happy to provide further pointers on how to do so, we are a bit reluctant to speculate on their effects regarding the occurrence of false positives. It seems unlikely that functions with very similar shapes would produce more false positives; however, sensitivity to functional forms is not unheard of, especially if the links are entirely deterministic.

Reviewer 02 May 2023 09:41

I think this is a very interesting paper, and it's nice to see that IPMs might indeed be a promising method for estimating species interactions. However, I have two main concerns with the paper in its current form.

First, I wish this paper was less dependent on Barraquand & Gimenez (2019) and stood a bit more on its own. As it is, most of the methods are described on the basis of the earlier paper, focusing on how they do or don't differ from that one and cannot be properly understood on their own. I would much rather read a thorough description of the methods used in the current paper, along with a biological justification and explanation, instead of reading that “this paper did the same as in that paper but with this small change”. This is a recurring topic, as almost every section in the methods starts with a reference to this earlier paper.

Authors response

We understand and have now reduced the number of references to the previous study where not needed (e.g., lines 79, 110, 150). However, we note that it is important—and not always sufficiently appreciated—to describe clearly which assumptions are similar or different from previous work, in order to disentangle why results might change.

Second, and more importantly, always fitting the same IPM as was used to generate the data (except for the one case where species interactions were included in the model and not in the data) does not seem like a realistic or useful test of using IPMs to estimate species interactions. Of course, it's a nice first step to isolate this one effect, but my worry would be that other variation in the data not being accounted for in

the model would cause spurious estimates of interactions. This could be observation error (which is assumed to be known in the current paper), or environmental variation causing stochasticity in other vital rates or intraspecific density dependence, for example. Even something as simple as fitting a model without allowing for random temporal noise to the data set where such noise is included. Without such considerations I find it hard to trust that these models would actually work in a real situation.

Authors response

Fitting the model used to simulate the data was required to assess the identifiability of model parameters (and assess the amount and type of data needed for practical identifiability) and to evaluate the coverage of parameter estimates (if the two models are different we already know that coverage will likely be non-nominal). In other words, any statistical evaluation should start by generating data under the fitted model.

That being said, we now encourage future work to explore the robustness of predator prey models to a range of credible model misspecifications (citing work from Plard et al 2021 using it for one-species IPMs) at the end of the discussion lines 294-303.

Minor comments:

This is just a small piece of advice: the first sentence of the introduction is a bit hard to follow and rather dry. A better first sentence might make the paper more attractive to the casual reader.

Authors response

We have now removed this sentence and split it into several simpler sentences lines 22-25:
“Estimating ecological interactions between and within species through models of their joint population dynamics is a task which requires large amounts of data. Indeed, with potentially as many as q^2 interaction parameters for q model compartments (combination of species and age classes), the number of parameters to estimate can climb very rapidly. Therefore, ecological statistics searches for improved ways to infer such population-level interaction strengths.”

L43-47: I'm not sure I follow this sentence and I don't understand how general these results are meant to be. What is the relation to the rest of the paragraph. In short: what message are you trying to get across to the reader in this sentence?

Authors response

We have now rephrased this sentence (lines 43-46) to make our message clearer: "Additionally, an empirical study in a bird predator-prey system (Queroue et al. 2021) was able to detect the expected bottom-up demographic linkages from prey to predator but not the expected top-down relationships, suggesting that those may be too weak to be detected."

Second paragraph in the introduction: It would be nice to link your study to studies like Mutshinda et al (2009) – Proc. Biol Sci 276 (2923-2929) where a large number of species are modeled and methods are used to identify non-zero interactions.

Authors response

Mutshinda et al. (2009) is now cited line 60.

L122-125: It would be better to give the explanations of the notation right after the first equation in which they are used.

Authors response

We have now done so, thank you.

Equations 6-17: Please give biological interpretations of these

Authors response

We have now done so lines 149-158.

L168: Do you really mean i.e. ("in other words") here?

Authors response

Thank you, we have replaced "i.e." with ":".

References

Auger-Méthé, M., Newman, K., Cole, D., Empacher, F., Gryba, R., King, A.A., Leos-Barajas, V., Mills Fleming, J., Nielsen, A., Petris, G. *et al.* (2021). A guide to state-space modeling of ecological time series.

Ecological Monographs, 91, e01470.

Besbeas, P., Freeman, S.N., Morgan, B.J. & Catchpole, E.A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58, 540–547.

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

Plard, F., Turek, D. & Schaub, M. (2021). Consequences of violating assumptions of integrated population models on parameter estimates. *Environmental and Ecological Statistics*, 28, 667–695.

Warton, D.I. (2018). Why you cannot transform your way out of trouble for small counts. *Biometrics*, 74, 362–368.