Review for:

Using integrated multispecies occupancy models to map co-occurrence between bottlenose dolphins and fisheries in the Gulf of Lion, French Mediterranean Sea

In this paper the authors extended the Rota et al. (2016) occupancy model for potentially interacting species to allow for the inclusion of multiple data sources to estimate the latent level of the model with increased precision. I liked the paper, and the need for this research is clear on couple fronts. First, the authors point out how important it is to map co-occurrence of dolphins and trawlers in the case study. Secondly, and perhaps more generally, the Rota et al. model is very data hungry. As such, being able to get more data from multiple data sets is a great selling point (and improvement). I would contend that the authors don't sell this second point enough, it gets brought up a bit in the discussion, but it could perhaps have a little more space delegated to that specific point.

I did have some questions about the mathematical exposition in the methods, as some bits were a little unclear. As I extended the Rota model out to a dynamic framework, I'm very familiar with the model so was able to fill some bits in. Others though, may struggle, and so some additional details would no doubt help others better understand and apply this modeling framework.

Following this, I have more specific comments for each section. I hope the authors find them useful. Happy to answer any specific question the authors have, my email is mfidino@lpzoo.org

- Mason Fidino

Abstract

Top-level thoughts

1. The abstract currently does not have any results of the analysis, is there a bit that you could sprinkle in there?

2. Should occupancy model or integrated species distribution model be part of the keywords?

Line by line comments

line 16: maybe 'estimate' instead of 'allow mapping?'

line 17-18: Better to say 'Here, we extended multispecies occupancy models by integrating multiple datasets' instead of 'Here, we illustrated how to extend....' Both get the same point across, but the former does so with fewer words.

Line 20: Could you get away with just saying 'data' instead of 'detections and non-detections?' I get what you mean, but it currently feels very specific for the abstract.

Line 20: 'both' is redundant.

Line 21: 'map' should be past tense (mapped their) to follow with the rest of the abstract and the current sentence

Line 21-23: The start of this sentence is a little unclear. I think you just need to change the start so it begins with 'Integrating several data sources...'

Introduction

Top-level thoughts

1. The end of the first paragraph introduces multispecies modeling, which comes off as a bit of a non sequitur. I think it would help to drop that bit on 'but not in multispecies modeling to our knowledge.' Instead, I think it would help to be more direct about what the previous approaches lack and therefore could be improved upon. For example, why is the shortcoming with using trawler data as a covariate in dolphin distribution models? Adding a final sentence here with something like 'While a useful starting point, such approaches {fill in reason why they are not perfect}.'

2. Paragraph two switches topics from mapping to occupancy models, I would consider either a) changing the topic sentence(s) so it is more occupancy model focused or, perhaps an easier approach would just be to b) make it two paragraphs.

Line by line comments

Line 28: interactions between whom? My guess is dolphins and fishers, but it would help to be specific.

Line 61-63: This last sentence feels more like the topic sentence for the following paragraph. This would also help to lengthen the final paragraph of the introduction, which is a little short.

Line 65: Maybe replace 'aiming to study' to 'to quantify'

Methods

Top-level thoughts

1. I'm not sure about the reasoning for not performing a deep investigation of ecological predictors. Does that mean you did not explore different covariates in your model? The model can be explained and then a set of models that represent different hypotheses could be fit. I have no issue here with the current approach, but I would perhaps try to explain why you did this in a different way (instead of relying on clarity). Perhaps, as the dolphin and trawling data represent one case study to demonstrate the use of this integrated model, you did not perform a deep investigation of ecological predictors. In practice, different models could be fit to that represent varying hypotheses and multi-model inference could be performed.

2. I'm guessing a softmax was used as the link function? I'd be specific about that.

3. The Rota model, as well as my extension of it, uses second-order parameters for the combined state (your state 4), which are associated to whether the species occur together. The equation you provide on line 112 lacks such a

parameterization, and instead it appears that a new set of parameters are
estimated for that specific state. Why was this done? To me, it makes it more
difficult to assess whether or not the species occur about as much as you would
expect and still requires the same number of parameters. A little bit of
explanation here would really help.

4. Was depth centered and scaled before the analysis?

5. The methods say there is more information about the smoothing function. I did
not see it in either of the supplementary info.

Line by line comments

Line 85: depth of what, and in what units? Also, why was depth included (providing a
little logic here will help the reader follow along with the ecological side of
things).

Line 108: What does the data look like for grid z? Is z a binary vector of length 4
that represents the latent state, where each element is associated to one of the four
model states in order? I'd provide an example for the reader what that may look like.
For example, if the dolphins and trawlers were in a given grid cell then z would be
`[0, 0, 0, 1].` Just a little something here to make people think about what the data
should be oriented as would help.

Line 112: What does /psi represent? it is not represented with the above subscripted
/psi values. You could be more specific about this in a couple ways. Generalize the
equation with a couple of extra subscripts so a_0, a_1, and your smoothing term have
an extra subscript for k in 1,...,3 states (excluding the no species present state)
such that it is

```
logit(/psi_k) = a_{0,k} + a_{1,k} depth + s(X,Y)_k
```

or just show state linear predictor as an example. However, see top-level thought 3
above as I have some uncertainty in how this model was parameterized if it follows the
Rota. et al. model.

Line 113: Aren't there parameters in the smoothing function as well that get
estimated?

line 129: Does z also have i and j subscripts?

Line 134-149: Seems like a great opportunity for a table here.

Line 152: Maybe add that the columns sum to 1? That fourth column is a lot to unpack,
but the component probabilities in there line up with the states on lines 134-149.

Line 153 - 155: I know exactly what you mean here, but the wording makes it difficult.
Here is my attempt, which could still use a little improvement:

```
Each observation y was linked to the ecological state z. To do so, let /theta_z
represent a row column of /theta that lines up with the latent state of a given grid
cell.
```

Line 162: to implement OUR GAM.

Line 166: Would it be better to provide a citation to zenodo? The GH repo can always get deleted / change, so it's not really the best thing to link to in a paper. And looking at the GH repo, some extra info in the readme would really help other people unpack this project. I sometimes go a little overboard with mine, but here is one example of a readme repo from a recent project. Providing adequate metadata and the like helps ensure that users have the appropriate background knowledge (or at least access to it) to ensure they understand your data.

https://github.com/mfidino/conflict

Results

Top-level thoughts

1. I know the paper is mostly about developing this model, but it would help to have a few more ecological bits peppered into the results. For example, what was the average space use (occupancy) for dolphins only, trawlers only, and both together?

2. Looking at figure 1.3 I am wondering about what the total amount of sampling effort that is possible in a 5 km by 5 km grid cell? I see that the effort is past 40 km. How much gets reliably 'covered' by 1 km of sampling, or is that something that could even be known (e.g,. probably depends on the person, weather conditions, etc.).

Line by line comments

Line 177: typo 'trawlers only or that trawlers only use.' Should also be 'a' grid cell, not 'the' grid cell.

Line 182: How was significance assessed?

Line 195: Could you add in one little bit about how much, on average, precision increased? Right now, this is a little vague and may not encourage someone to jump into the supplemental material. The increased precision you often get with integrated models is one of their main selling points!

Discussion

Top-level thoughts

1. Do you have any guidance about things to consider when locating datasets for integration (to be used with this model)? You have some other great examples of other data sources, but what kinds of data should people be looking for? Likewise, what if the detection processes were wildly different from one another such that it would be difficult to squeeze them into /theta (e.g., sampling in different locations)?

2. The discussion is a little thin here. Some other bits that you may want to consider adding would be a) caution when interpreting co-occurrence parameters b) any other additional caveats about this modeling framework that would modify model implementation or interpretation c) how this described approach differs from the aforementioned mitigation techniques d) more generally (outside of

your case study) what are the applications of this model? e) Not only were you able to combine datasets, this framework lets you model interactions as a function of covariates. That is definitely a selling point & something important for your case study (look at how much co-occurrence varied as a function of depth). f) any suggestions for further model extensions? For example, combining datasets with dramatically different sampling processes?

Line by line comments

Line 198: Makes sense to have showcase be past tense.

Line 198 - 205: For interactions to take place they have to co-occur, so knowing where that is most likely is definitely the first step. From there, more pointed surveys / studies could be done to determine what features move you from co-occurrence to conflict.

Tables & figures

Top-level thoughts

Fig 1. Adding the observed data is great here. When this figure is made a little smaller I fear that the axis text is going to be too small (both for the titles and axis text). Making them a bit bigger (and black instead of gray) would help. Why is the y axis here occupancy when in figure 2 the legends are space use?

Fig 2. You should be explicit that some of these subfigures are marginal occupancy probabilities (and probably put somewhere in the methods how to calculate these). Also, are the marginal occupancy probabilities what you really want to show here? Not much additional information is gained it seems (looks very similar to the cooccurence plots so I'm not really sure what is gained from them). Regardless, great color schemes here!