PCI – Multispecies occupancy

Authors' answer: In this document, we included the copy of each reviewer report and our answer to each of their comment.

REVIEWER #1

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
```

Authors' answer: Many thanks for your comments on our manuscript. You're right that we have two selling points in this manuscript, i) the extension of multispecies occupancy models to integrate multiple datasets, and ii) the illustration of the multispecies occupancy framework to study marine mammals – fisheries interactions. Following your comments and those of the second reviewer, we acknowledge that the modelling point is more valuable and we put more emphasis on it in the revised version, presenting the dolphins – trawlers situation as a case study. In the following lines, you will find our answer to each of your comments. Thanks again for your help in clarifying and improving our work.

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?

Authors' answer: In the revised version of the abstract, we added that we assess the performance of integrated vs single-dataset multispecies occupancy models, and that our approach enables to map co-occurrence probability as a function of environmental covariate although we did not explore in details the ecological processes.

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

Authors' answer: We added "occupancy models" and "integrated models" as keywords.

Line by line comments

line 16: maybe 'estimate' instead of 'allow mapping?'
Authors' answer: Done.
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.
Authors' answer: Done.
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.
Authors' answer: Done.
Line 20: 'both' is redundant.
Authors' answer: We corrected that.
Line 21: 'map' should be past tense (mapped their) to follow with the rest of the abstract and the current sentence
Authors' answer: Done.
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...'
Authors' answer: We rewrote this sentence.

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}.'

Authors' answer: We dropped the section you suggested. Instead, we added precisions at lines 45-50 that acoustic behavioral observations of interactions are crucial but spatially limited, and we also pointed out that adding a covariate in distribution models does not take into account biotic interactions and might lead to biased estimates of species occupancy.

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.

Authors' answer: As you suggested, we made it two paragraphs. The first one lines 51-64 to describes why the challenges occurring when mapping species interaction. The second one lines 65-72 presents multispecies occupancy and why it is relevant to integrated datasets.

Line by line comments

Line 28: interactions between whom? My guess is dolphins and fishers, but it would help to be specific.
Authors' answer: You're right, we detailed it in the revised version.
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.
Authors' answer: In the revised version, we lengthen the last paragraph as you suggested. We added that we also performed the comparison between integrated vs single-dataset multispecies occupancy models. We also added a sentence about the data we used lines 76-81.

Line 65: Maybe replace 'aiming to study' to 'to quantify'
Authors' answer: Done.

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.

    Authors' answer: You're right. We intend to illustrate the extension of integrated multispecies occupancy model, to assess the benefit of data integration, and to suggest that multispecies modelling can be interesting for studying dolphin - fisheries interactions. Concerning the ecological exploration of our case study, the restricted size of our dataset limits the possibility in terms of the number of covariates that could be included to the model. But, you're right, we can totally run competing model and perform model comparison. We tried to better justify this choice lines 104-109.

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

    Authors' answer: We added line 140 that we use a generalized logit link to ensure that the probabilities sum to 1.

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.

    Authors' answer: Sorry for not being clear, we also used second-order parameters, you can see R-codes supplementary materials. We corrected formulation in the revised version lines 143-152.

4. Was depth centered and scaled before the analysis?

    Authors' answer: Yes. We added precision line 103.

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

Authors' answer: We provide details about the smoothing function in the Supplementary Infos about Rcodes.

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).

Authors' answer: In the revised version, we precise that we refer to the depth of seabed in meters. We added that we scaled depth value before using it in the models, lines 103-104.

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.

Authors' answer: We added an example lines 127-130, of the possible z values saying that "the latent occupancy state can take 4 values: "z=[1,0,0,0] if neither dolphins nor trawlers use the grid-cell, z=[0,1,0,0] if dolphins use the grid-cell but trawlers do not, z=[0,0,1,0] if trawlers use the grid-cell but dolphins do not, and z=[0,0,0,1] if both dolphins and trawlers use the grid-cell. Then, ignoring the grid-cell index, our multispecies occupancy model estimated 4 occupancy probabilities."

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.

Authors' answer: We detailed the Psi equation with the k subscript as you suggested lines 148-149. We also precise the i subscript for each grid-cell.

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

Authors' answer: Yes, we precise it in the revised version, lines 150.

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

Authors' answer: No, as we only consider a static occupancy model z only has a i subscript related to grid-cells but no j subscript related to sampling occasions.

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

Authors' answer: We were also tempted to add a table here. However, we do not clearly see how can we build such table that would not be enormous. While column can be the four latent occupancy states, we do not see any easy way to link them to sixteen observation events. Classical Hidden Markov Model matrix representation between states and observation leads to the large \theta matrix we already displayed.

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.

Authors' answer: Using comments from both reviewers, we added more details to describe the big \theta matrix, including the fact that "each column of $t(\theta)$ represent the all observation probabilities for the latent state of a given grid-cell and hence sum to 1." Lines 181-184

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."

Authors' answer: We tried to rewrite the sentence with your suggestion, it is now lines 183-184.

Line 162: to implement OUR GAM.
Authors' answer: Done.

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
Authors' answer: We already created the Zenodo and you're right, we now cite the Zenodo repository, (https://zenodo.org/record/7871492)

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?

Authors' answer: We included marginal average space-use probabilities in the results and detailed the interpretation in terms overlapping distribution lines 221-225. More important, following your suggestions and those of the other reviewer, we added the comparisons between integrated vs. single-datasets and multi- vs single-species occupancy models in the results sections, lines 241-256.

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.).

Authors' answer: You're raising an important question, and you answered part of it. It depends on the sampling method. Area covered by aerial surveys depends on the flight altitude and detection will depend on the flight speed, weather condition, and observer experience. For boat surveys, covered area will depend on the elevation of the observation platform above sea level, while the same variables as for aerial sampling will affect detection.

Line by line comments

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

Line 182: How was significance assessed?
Authors' answer: In the revised version, we precise line 206-206 that we considered an effect of covariate as significant when its 80% Credible Interval does not overlap 0.

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

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?

Authors' answer: We corrected both the axis size and changed the y title for "Space-use probability".

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 co-occurence plots so I'm not really sure what is gained from them). Regardless, great color schemes here!

Authors' answer: We thought that the lower panel of Figure 2 would help to see the spatial distribution of raw data. But following comments by the other reviewer, we included a data figure in the Methods section and we replaced the figure of the marginal probabilities by the plots of the comparison between integrated and single-method occupancy models.

REVIEWER #2

Summary:

This manuscript describes an original method for integrating different datasets into a multispecies occupancy model, and showcases this modeling framework using the case study of dolphin-trawling interactions in the Mediterranean Sea (hence making the analogy between trawlers and another species). More specifically, the authors developed an integrated Bayesian model (NIMBLE-based; code available online) which, fitted to detection/non-detection data, estimates how environment affects where trawling activites and dolphins occur and co-occur while implicitly accounting for the potential impact of each component on the distribution of the other. Taking advantage of the Bayesian approach, the multispecies co-occupancy model makes use of all available information by making very different surveys (onboard and aerial) comparable through an estimate of their respective sampling efficiency. Using the multispecies occupancy model in a predictive way, the authors can then map the co-occurrence of dolphins and actively fishing trawlers in the Gulf of Lion, highlighting the potential of the approach in the context of by-catch management.

Main comments:

I appreciated reading this manuscript, which is pretty straightforward and clear, well illustrated, and goes with helpful complementary information and documented code. The ecological modeling work described in the manuscript is of great quality. The authors use the appropriate tools to answer the main ecological questions and show a fine understanding of the assumptions underlying each step of the modeling procedure. The framework they developed is transferable to other case studies and contexts and may have concrete applications in the near future for assessing and managing fishing-wildlife interactions.

The main asset of this work is the modeling framework itself, both with respect to (i) the way it maximizes the amount of information that can be incorporated and (ii) the originality of its application to fishing-protected species interactions by considering the trawlers as a second "species" of the occupancy model to account for the potential interdependence of their distribution. As partially highlighted by the authors in their Introduction, to date, many studies mapping the distribution of interactions between fishing and protected species rely on data biased towards one of the components involved in these interactions that are sometimes combined together and rarely in an adequate way: reported by-catch, observers onboard fishing boats, or simply fishing effort distribution for the former, megafauna-focused survey for the latter. Although surveys and monitoring programs able to provide information on both megafauna and fishing remain limited nowadays, we can expect that the increasing use of new autonomous monitoring systems (e.g., drones or gliders) and the development of tools for the analyses of the data they collect (~ automatic analyses through machine learning) will favour the acquisition of such data in the near future. In that context, the present modeling approach looks particularly promising.

On a minor and purely methodological aspect, I also liked how the model was assembled. To couple the occupancy and the detection modules of the model, the authors hacked a very popular tool in marine ecology for building species distribution modeling: generalized additive models (and associated mgcv package). While Bayesian integrated models are less accessible and mastered by a smaller part of the community, the framework proposed in this article and the available supplementary material might help researchers transition or get more familiar with such types of frameworks.

The results of this model application are slightly frustrating as it seems that (I'm putting it provocatively), the conclusion is that "dolphins and fishing boats co-occur where they occur": the

spatial patterns in the probability of occurrence of dolphins/trawlers are very close and are similar to that of their co-occurrence. It suggests that, probably due to some modeling choices, data constraints, and the reality of fishing- dolphin interactions at that scale of time and space, interactions between dolphins and trawling activity are poorly structuring their distribution in comparison with environment. Nonetheless, this case study remains helpful in showcasing the potential of this approach.

For all these reasons, I think this work is definitely valuable for the ecological modeling community, with perspectives for management applications. Once that said, I have some reservations regarding several points.

Authors' answer: Thank you very much for your work on our manuscript. Your comments helped us to have a clearer view of the most important strengths and weaknesses of the manuscript. In the revised version, we added in the main text the comparisons we made regarding integrated vs single-dataset models, and comparison you suggested to assess the benefit of the multispecies framework. Due to limited length of the manuscript, we do not perform in depth discussion of these modelling results but we detailed them in two annexes. We hope you will like the revised version. Many thanks again!

My main reservation is related to several modeling choices made in the present study that would require a better justification and whose implications regarding the outcomes of the study should be discussed. A first example is the temporal and spatial resolution of the data and model. I would expect that the manuscript describes more basic choices regarding, for instance, grid-cell size or the seasons covered. All these choices are probably driven by the amount of data available but need to be justified. The data description in the present version of the manuscript does not help: there is no information about the temporal coverage, and information on the spatial coverage is only provided in the Annexes.

Authors' answer: You are right that our limited dataset precluded us from using a number of covariates and constrained us on the resolution of the grid we used. We added precision about it lines 104-109 saying that "At the date of our modelling developments, the resolution of the grid and our capacity to explore multiple environmental descriptors of co-occurrence patterns is impaired by the limited size of our datasets."

Maybe interspecies interaction effects on co-occurrence may be stronger with a finer resolution..? All of this should be discussed. Another example is the choice of the covariates included in the distribution model.

Authors' answer: In the Methods section of the revised version, we justified that we could not explore a multitude of ecological covariate due to the limited amount of data in our dataset. We included i) depth that is established as a relevant covariate to model the ecology of marine species, and ii) a GAM on geographical coordinates that permit to model spatial variation in details despite a geographical GAM did not allow to understand the ecological processes linking marine species and their environment. As more data is currently being collected on that case study, we acknowledge that future models would strongly benefit from further exploration of ecological variables.

To explain the presence or absence of dolphins and fishing, the authors chose one environmental covariate only (and a spatial tensor), i.e. depth. I know that even complex habitat models built for odontocetes often identify depth as one of the primary variables structuring their distribution. Still, this should at least be justified with some references.

Authors' answer: You're right, as we pointed out in our previous comment, we added more details about covariate, references, and modelling choices lines 101-103, and 105-109.

Also, due to technical limitations or fuel coats, most of the trawling activity is generally restricted to the continental shelf, 0-250m, and in smaller extent, the continental slope (trawlers are operating deep in this study but it remains difficult for me to analyze since the "trawlers" are not defined in the M&M; see specific comments).

Authors' answer: We added details lines 96-98 that we defined trawlers "every commercial fishing boat that we observed actively dragging."

As a consequence, integrating only depth could limit the pertinence of the model. It will likely show a depth effect at the scale of the whole study area, which includes deep waters not frequented by fishing, while interesting patterns in shallower waters, i.e., where most of the fishing activity occurs, may be missed; Especially since depth has a linear effect.

Authors' answer: We could not add multiple covariates due to limited data. Although we could consider quadratic effect on depth, bottlenose dolphins are known to have a preference for shallow waters (Labach et al., 2021; Torres et al., 2008). We explained and acknowledged the limitations lines 105-109 of the data section.

Another important reservation is about some choices the authors made regarding the manuscript's content. A complementary work was conducted to characterize how the integrative aspect of that approach was crucial. One of the annex documents compares the outputs of the models when integrating one dataset only and both datasets. I find this part relatively interesting and, given (i) the brevity of the manuscript and (ii) the number of statements regarding the interest of integrative approaches, I think it would make sense to move it into the main manuscript. It is just a suggestion, though.

Authors' answer: As you suggested, we included our comparison of integrated vs single-dataset occupancy models in our manuscript. We wrote new paragraphs in the Methods lines 186-200, Results lines 241-256, and Discussion sections lines 266-272 accordingly.

More importantly, the Introduction emphasized the importance of accounting for the role of interspecies when studying their distributions. As far as I understand, this aspect motivated the authors to adopt the multispecies co-occurrence modeling approach. However, no output, figure or analysis allows us to assess the benefits of this approach. I think the authors should find a way to highlight it better (this is, to me, even more important than discussing the value of data integration). From what we can see of the model outputs, the advantage of the co-occupancy model over 2 separate monospecies approaches may be relatively limited in the present case study. I am not asking the authors to run the model in a monospecies configuration (unless it is straightforward!). Nonetheless, the manuscript would benefit from an assessment (even qualitative) of the added value of the multispecies aspect. At least, I would expect much more about this topic in the Discussion.

Authors' answer: You pointed out an interesting point despite being hard to evaluate in detail. To illustrate the point you made, we run each analysis separately. We fitted i) a dolphin integrated occupancy model, ii) a trawler integrated occupancy model, and iii) the multispecies integrated occupancy model. Our results suggest that the bigger gain for multispecies framework is in terms of associated precision of the ecological estimates. We wrote a Supplementary file describing these results and we included one figure in the main text. See our comment above.

Last, I strongly recommend the authors to carefully check their text as I found a lot of typos should not be so numerous at that stage (especially those detectable with Word or other automatic correctors).

Some sentences were not clear in different places in the manuscript; I tried to flag them in my specific comments.

Authors' answer: We double checked our reading to correct for as many typos as possible. Sorry about the multiple ones that occurred in the previous version.

Note that I reviewed the code provided by the authors but didn't take the time to run the model by myself.

Conclusion

For the reasons mentioned above, I consider the present manuscript highly valuable and worth publishing. However, I suggest it is revised to tackle the issues mentioned in my reservations before any acceptance.

Authors' answer: Thank you !

Specific comments:

Summary

L22 – *The* possibility (ability?)

Authors' answer: We wrote a new sentence, hence this word no longer exists in our abstract.

Intro

L.27 – the fact that it is especially affected by anthropogenic [not "anthropic"] is not only resulting from its "seaway function" – reformulate

Authors' answer: Done.

L.29 – reference for "increasing interactions"?

Authors' answer: We refer to Avila et al., (2018) we change the wording to be more cautious than "increasing", lines 32-33.

L31 – there are probably some experts in the co-authors, but in my mind, "depredation" is the fact that these predators feed on fish caught (or at least localized and targeted) by fishers. So "foraging behind trawlers" IS predation (leads to depredation), right? Then, OK, depredation behavior can occasion by- catch.

Authors' answer: Yes, you're right! Depredation behavior includes predation on fish stocks localized or being fished by trawlers. It could occasionally lead to by-catch.

L34 – "depredation pressure" is hard to quantify. Maybe mentioning "depredating behavior" is sufficient..?

L37 – "in multiple locations worldwide"◊"worldwide"?

L38 - "Following mortality events that have been documented about bottlenose dolphins (Manlik et al. 2022), interactions have raised conservation concerns and mitigation measures trialed thus far have not proven effective (Snape et al. 2018, Bonizzoni et al. 2020)" ◊ "Following documented bottlenose dolphins mortality events (Manlik et al. 2022), interactions have raised conservation concerns and mitigation measures tested to date have not proven effective (Snape et al. 2018, Bonizzoni et al. 2020)"

L41-44 – Bunch of different approaches listed here; try to be more specific and quickly display the advantages and drawbacks of each method.

L47 – be more specific (this isn't the first sentence of the intro); try something like "….conservation conflicts and is, therefore, particularly strategic in the case of the mammal by-catch issue"?

L51 – "needs"

L47-49 – Maybe I would put just after this sentence the reference to Pirotta (L43): it is step 0 of interactions mapping. It would support your statement about the need for multisp co-occurrence model.

L58 – "underestimation of species distribution" what do the authors mean?

L62 – "to combine multiple datasets into"◊"combining multiple datasets within"?

M&M

L72 – Data paragraph: maybe the authors could add a short description of the study area to highlight its relevance regarding the dolphin-fisheries interactions

Authors' answer: We added a figure of the detections data and line transects following your comment, line 110.

L76 – maybe first the megafauna (main focus), then human activities?

Authors' answer: Done.

L77 – "detections and non-detections data"◊"detection /non-detection data"?

Authors' answer: Done.

L80 – Onboard survey?

Authors' answer: Done.

L82 – Ony detections here? Or non-detections too?

Authors' answer: Corrected.

L83 – harbours

Authors' answer: Done

L83 – Trawlers: which type (all? Pelagic? Demersal? Which size?)? We want to be sure that it is pertinent with interactions that have been observed in the literature

Authors' answer: We now write that observers considered every commercial fishing trawlers actively dragging, lines 96-98.

L86 – "as the transect length (in km) of each monitoring program for each grid-cell"◊"as the total length (in km) of transects conducted in each grid-cell by each monitoring program"?

Authors' answer: Done

L87 – Using depth as the only covariate should be, to my mind, justified.

Authors' answer: We added justifications about that lines 101-104. See our previous answers to this point.

L81-87 – We have here information about the spatial coverage of the dataset – but we don't have anything about the temporal resolution/coverage.

Authors' answer: We added the year and the temporal coverage in the data paragraph lines 88, 90-92 of the data section..

L89-97 – ok, helpful clarification

L112 – I guess X and Y are long/lat of the center of the grid-cells?

Authors' answer: Yes, we added precisions about it lines 141-142.

L114 – Specify the family chosen + the type of smoother

Authors' answer: We specified that we used penalized splines in the Rcodes supplementary material.

L114 – Based on the study area map, I have the feeling that correlation may be strong between the covariate, depth, and the spatial smoother, e.g., X positively correlated with depth, and Y negatively correlated with depth. I guess the authors have already checked this but could they discuss this in this paragraph? Is there a correlation, is it strong etc... ? If there is a substantial correlation, what are the implications for the inferred probabilities of co-occurrence?

Authors' answer: We tested for correlation between X, Y, and depth. We did not consider that our covariates are highly correlated as |r|<0.7. We also consider that the problem of including highly correlated covariate relies in the interpretation of the covariate effect. As we mainly care about making accurate predictions of occupancy probabilities rather than environmental interpretation (i.e. hard to interpret a geographical GAM), including highly correlated variables would not significantly matter (Sonderegger, 2023).

L119 – "winter, spring, summer, and autumn" – first time that authors mention seasons, should be described in the Data part.

Authors' answer: As you suggested previously, we detailed the temporal scope of the datasets in the Data section and introduce these seasons.

L136-157 – I find this part a bit laborious. I am actually not sure that the 16 feet long list of "observation events" is helpful here. Maybe it would be ok to transfer it to an Annexe, together with a more detail explanation of how the theta matrix is built; and replace L135-154 with smthg like "These four probabilities could then be used to explain the simultaneous detection or undetection of each species by each survey, consisting 16 observation "events" (= (2 species) ^ (2 detection status) ^ (2 surveys))." I let the authors think about it.

Authors' answer: Following your advice, we removed the 16 feet long list in the supplementary material, and we added as you suggested the 2 ^2 ^2 sentence, lines 175-176.

L159 – I would instead write "multiple datasets/data types integration."

Authors' answer: Done.

L166 – 80% is not consistent with what is indicated in the caption ◊ which one is correct?

Authors' answer: Thank you for seeing that. The 80% Ci is right, we corrected the caption.

Results

L169 – Would it be possible to compare the relative importance of the depth effect vs spatial tensor? The most important results are displayed, but people may be curious about it.

Authors' answer: Yes, it would be possible either performing model comparison between models with and without the GAM, or either by assessing the covariate effect size. However, we do not discuss this in our manuscript.

L171 – keep the greek letter for psi

L171 – detail but help clarify the paper: dolphins on Fig 1.3 should be in green to respect the color choices of Fig1.2 (if the authors choose to include figures from the Annexe in the main, they should also check this aspect)

Authors' answer: You're right. We checked that the color code remains consistent throughout the manuscript, and changed color theme of figure 1, 3, and 5.

L177-178 – May also be put in the M&M: data limitation issue should be mentioned prior modeling

Authors' answer: Yes, in the revised we added that at the end of the data section, lines 105-109.

L179 – "that trawlers only or that trawlers only" dolphins?

Authors' answer: Corrected.

L182 – "Although, dolphins"

Authors' answer: Corrected.

L186 – Just a visual thing: proximity between the 2 upper and the 2 lower panels is confusing when reading the figure. Maybe consider putting more space between the titles of lower panels and upper panels.

Authors' answer: Done.

Discussion

L203-207 – One point that is worth commenting on is also the impact of the grid size that has been selected.

Authors' answer: We did not discuss it in depth but we acknowledged this is a limitation due to limited in the data section of the M&M, lines 105-109.

L213 – "Supplementary materials": I would not be shocked to read more about the outcomes of this analysis in the Discussion.

Authors' answer: Following your advice, we included this results in the manuscript and add precision about them in Results and Discussion section lines 241-256 and lines 267-272.

L221 – campaigns◊surveys

Authors' answer: Done.

Annexes

In the script: please define all variables, even the intermediary variables – can be done in an intro to your code. Quickly showing the structure/content of the input data would also help.

Authors' answer: Done. In the R codes PDF file, we added two paragraphs detailing all datasets and variables used in the model.

Supplementary Material Annexe 2 overlaps a lot with the M&M paragraph of the manuscript – except for the data part, which is so short that it can be transferred into the M&M.

Authors' answer: Done. In the revised version, we added some part of the "Data" section of Annexe 2, into the main text of the manuscript, lines 100-113. On only kept the previous format of Annexe 2 as an explanatory document in the GitHub repository.

References

Avila, I. C., Kaschner, K., & Dormann, C. F. (2018). Current global risks to marine mammals: Taking stock of the threats. *Biological Conservation*, *221*, 44–58. https://doi.org/10.1016/j.biocon.2018.02.021

Labach, H., Azzinari, C., Barbier, M., Cesarini, C., Daniel, B., David, L., Dhermain, F., Di-Méglio, N., Guichard, B., Jourdan, J., Lauret, V., Robert, N., Roul, M., Tomasi, N., & Gimenez, O. (2021). Distribution and abundance of common bottlenose dolphin (Tursiops truncatus) over the French Mediterranean continental shelf. *Marine Mammal Science*, *n/a*(n/a). https://doi.org/10.1111/mms.12874

MacKenzie, D. I. (Ed.). (2006). *Occupancy estimation and modeling: Inferring patterns and dynamics of species*. Elsevier.

Queiros, Q., Fromentin, J., Astruc, G., Bauer, R., & Saraux, C. (2018). Dolphin predation pressure on pelagic and demersal fish in the northwestern Mediterranean Sea. *Marine Ecology Progress Series*, *603*, 13–27. https://doi.org/10.3354/meps12672

Sonderegger, D. L. (n.d.). *Statistical Methods II*. Retrieved 25 April 2023, from https://bookdown.org/dereksonderegger/571/

Torres, L. G., Read, A. J., & Halpin, P. (2008). Fine-scale habitat modeling of a top marine predator: Do prey data improve predictive capacity. *Ecological Applications*, *18*(7), 1702–1717. https://doi.org/10.1890/07-1455.1