Editor Comments

New revision requested

Dear authors,

We have now received three reviews of your revised manuscript.

Two of them are positive and request only some minor revision, but the third still raises some important points that need to be carefully addressed in a new revision.

Thank you again for submitting your work to PCI Ecology.
We look forward to receiving a new version of the manuscript.

Best wishes,

François

Dear recommender and reviewers,

We thank you for your follow up comments. We answered all of them below (in blue) and amended our article to include your suggestions.

Again, changes can be seen here: https://draftable.com/compare/ZFlzGUQfsTxw

The most significant updates are additions to the discussion, as suggested by Jean-Olivier Irisson. In particular, we tried to address his concerns about the impact of the two geographical predictors and the fact that our maps depict *relative* probabilities.

We hope that we answered all questions and suggestions properly and that you find this version of the article worth publishing.

Sincerely,

Gaétan Morand (corresponding author)

Review #1

Second review for "Predicting species distributions in the open oceans with convolutional neural networks"

General remarks

My main remarks in the first round of review were:

1. the need for a baseline model to compare the CNN with
2. a discussion regarding the goal of this model and whether it was a "niche" model
3. the sometimes surprising choice of species
4. the fact that the model was a classifier, not a regression and that it could have unexpected consequences
5. the inclusion of geographical predictors (hemisphere etc.) which led to artefacts in the predicted distributions
6. the caricatural aspect of the 2ºC increase simulation

I understand that the authors are not in a position to re-run the analyses presented in the paper. This makes it impossible to fully answer some of my remarks or those of other reviewers. They made a significant effort to review what they could and modify the text. I leave it to the recommender to decide whether this is enough to warrant a recommendation in the end.

Regarding my main remarks:

1.
The addition of the punctual DNN is welcome and clearly shows the advantage of using the spatial information. This is a good addition to the paper. Thank you.

2.
This is now better presented in the introduction and the word "niche" is not used anymore in the context of this paper, which is appropriate.
The discussion of the fit with theoretical distributions (in section 3.2) is now more cautious.

3.
The selection of species could not be changed. The introduction was altered to mention that the species selected were expected to have different distribution characteristics.
Most are mobile species, for which the dynamic SDM approach makes sense. I still think including the Acropora coral, the only sessile species in the list, does not make sense, in particular in light of the following remark.

We agree that the presence of Acropora can be surprising in the context of dynamic SDMs. But even though the genus is sessile, the environmental conditions do vary and may be more or less favorable to its presence: its ability to settle somewhere and to survive depend on the conditions. The inclusion of Acropora could be useful to future studies, for example to examine the effects of El Nino Southern Oscillation or to model the effect of ocean warming.

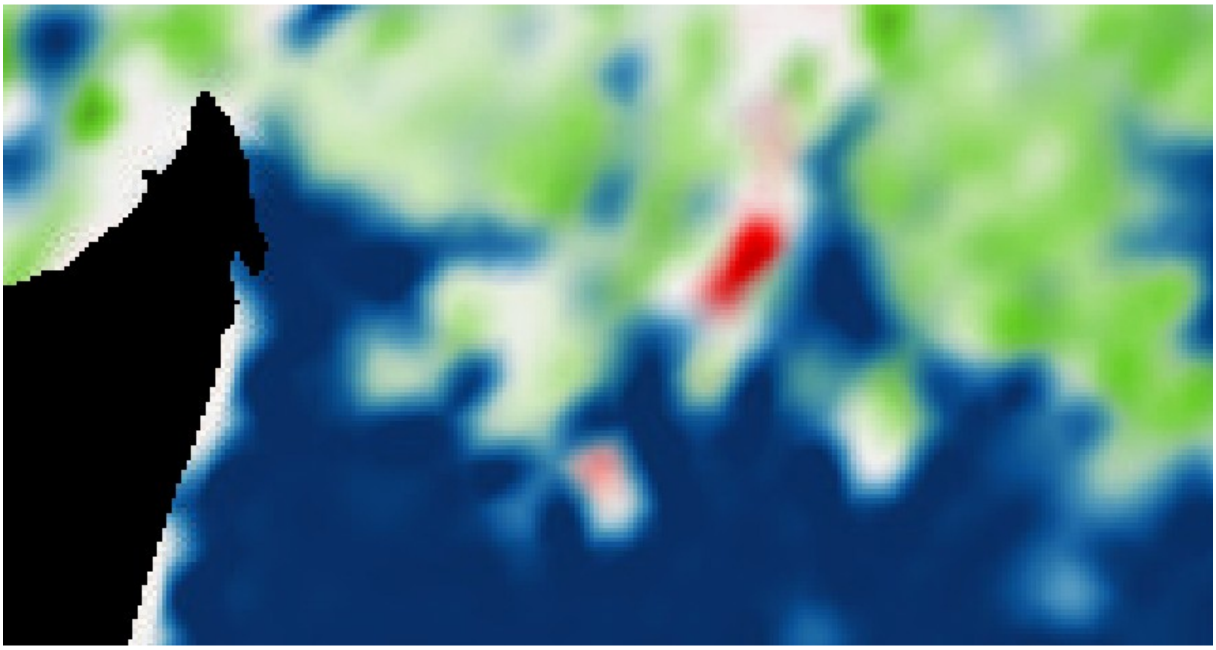
We added this into the article (lines 122-125).

4.
The nature of the model (i.e. a classifier) is now better presented, in a paragraph starting at line

206. A short paragraph in the discussion (line 345 and following) helps with the interpretation, but I think it is not enough, because the maps are actually very different from usual SDMs.

What is represented by each map is the *proportion* of each species among the 38 selected (assuming no sampling bias etc. as the authors point out).
This means the following, assuming a model with two species, A and B:
- in a pixel in which the conditions are maximally favourable for species A but absolutely not for species B, the value predicted for species A will be 1
- in a pixel in which the conditions are barely correct for A but still not favourable for B, the prediction will still be 1
- in a pixel in which the conditions are equally favourable for both, the predicted value will be 0.5 (as stated line 220)

This makes it very difficult to interpret a single map in isolation from others. It also explains some of the very spotty aspect of the maps: when conditions are still favourable for species A but also become favourable for another species, the predicted value for A drops. This is exemplified by the following image where blue is for Caretta caretta, green is Carcharhinus falciformis and red is Carcharinus albimarginatus; while the whole region may be favourable for the green turtle, the extra favourability of some pixels for the sharks make the habitat suitability for the turtle drop. This is purely a numerical artefact, not an ecological feature: these sharks only rarely eat green turtles. This is not wrong per se, but it will likely be wrongly interpreted by an unprepared reader (I know I was mistaken during my first review and I only realised this now because we have been working with proportions too and scratching our heads trying to interpret the maps).



So, overall, it should be made much more obvious that all maps are maps of proportions among the 38 species (more frequently and strongly than the mention at line 246) and their discussion should be made in light of that fact. For example, the absence of a species within its distribution range (section 3.3) may be caused by differences in season, immediate vs. long term conditions, etc. as discussed, but also by the fact that another species dominates in one part of the range of the target species. Do not hesitate to repeat yourselves; it is quite difficult to wrap one's head around the interpretation of maps of proportions.

We clarified that we are dealing with proportions / relative probabilities (lines 286, 292 and 298, in addition to previously written lines 112, 218, 251-253 and 357) and that this may cause artefacts (lines 315-317, 457-465 (new subsection))

Another way to present this is to say that your model is not a model of species distribution but rather a model of *community composition*: you are trying to define the proportion of each target species in each pixel and, in particular, which one dominates.

We find that the “community composition term” might be interpreted as more misleading because it suggests exhaustivity among all species of the community, yet we have only 38.

In terms of representation, to make this obvious, with few species, you could plot a single map with a colour in each pixel, resulting from a mix of colours proportional to the probability of each species; with three species, you could take Red, Green and Blue and the resulting RGB colour would inform on which species dominates. With 38 species, I am not sure what to do, but maybe you can restrict yourselves to a few?

We added a figure such as the one you suggest to the discussion to show this effect, but with two species only.

In the same line of thought, we tried to do a 3-dimensional PCA to display habitats in a single RGB figure, but the 38 species were too different for it depict any valuable information.

In light of this fact, the selection of Acropora makes even less sense: all others share similar habitats, some may compete for the same food source, etc. It makes more or less sense to consider them as a community and study their relative abundance; this is not the case for Acropora.
It could have had little effect on the results since it is a very coastal species while the others are all pelagic (therefore their distributions are not "competing"), but that supposes that the coastal conditions are different enough from the pelagic ones, which is not true given its distribution map (and your discussion at line 415). This is briefly discussed in the new version of the paper but is a serious limitation for the predictions in the inter-tropical Pacific.

There is no notion of abundance or competition In this study: the objective is to model habitat suitability, or species presence at best.
The only potential issue with Acropora is that it does not respond quickly to changing environmental conditions, which may hinder the model’s ability to isolate suitable conditions at high temporal resolution. But that is also true for some mobile but sedentary species such as *Cheilinus undulatus*. We added this dichotomy to the discussion. (lines 416-419)

As you note, a way to circumvent this "relative abundance" problem is to predict many species at once: this was, each species is coexisting with, probably, many others in each pixel (i.e. each set of conditions), the predicted values do not approach 1 but their value is closer to the true habitat of the species. But the example above shows that 38 is not enough here.

We added that 38 is not enough to prevent one species from dominating and avoid artificially lowering probabilities for other species. (line 465)

5.
This is not presented as a limitation of the study. I am not convinced by the solution ("blurring" the limits rather than having hard ones): there will still be a fake boundary, which makes no sense to animals around the equator for example.
I was looking for papers discussing the inclusion of such specific geographic predictors in SDMs but could not find a definitive paper. This search probably needs to be deepened and the discussion of the consequence a bit more advanced (in the absence of the possibility to re-run the model).

Ideally the equatorial barrier would only influence polar species and ocean barriers would only influence low latitude species: For each species only barriers outside of its range should have a significant impact. We expanded on that in the discussion and also propose two other solutions. (lines 444-453)

6.
The simulation was removed, which I think was the right decision.

Detailed remarks

56: While simple averaging indeed erases dynamic structures, it is still possible to use another summary in climatologies. For example, to predict the most favourable habitat of tuna, a climatology of the frequency of presence of fronts (from FSLE snapshots) is likely to be a relevant information. So add "If one uses simple averaging, the use of climatological..." and possibly rephrase this sentence a bit.

We added a sentence to mention the existence of this type of products (lines 55-57)

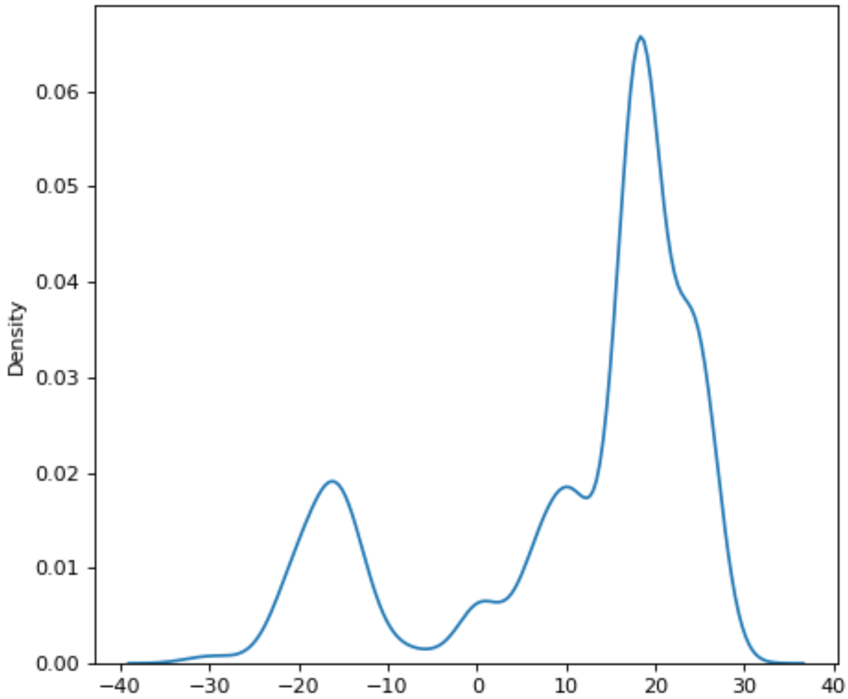
88: CNNs can be used for regression also, there is nothing specific to classification. I understand you added this sentence to highlight that *yours* is a classifier, but the current sentence is not specific enough.

We changed the sentence. (line 90)

133: Cleaning up species records was advised by another reviewer; you explained that it was impossible to perform all manually. I agree but some automated solutions could at least

allow you to automatically spot outliers in geographical/environmental space (e.g. kernel density estimation).
Seeing some of the initial point distributions, this would be worthwhile, with Acropora records in Southampton for example...

There are indeed outliers but they are too few to influence our results. As an example, below is the distribution of latitudes of Acropora observations in our training dataset. But we do agree that some cleaning could be beneficial (we added this line 139).



Discussion: while I appreciate the effort to include many remarks from the reviewers in the discussion, they often only amount to an acknowledgement of a problem and a statement that this needs to be further explored. I would have preferred a bit more depth, in discussion in general

The discussion now goes more in depth thanks to the latest additions in response to your comments. Unfortunately in many cases the fact that we cannot rerun the model training prevents us from conducting more comparative analyses.

PS: draftable.com was super useful. Thank you for using it!

Review #2

Predicting species distributions in the open ocean with convolutional neural networks. version 2

General comment on the revised version

During a first round of peer reviewing, and in addition to my review, the authors have received very thorough and high-quality reviews from the editor and from two other colleagues on the first version of the manuscript.

In this revised version, the authors did not rerun their model, nor changed their methodology, as this would have taken a lot of time, which several of them do not have because they are under short term contracts. While I totally understand this practical reason, I find that their answers to the reviews that they have received (in particular the very complete and constructive review of JO Irisson) are not as much detailed as the reviews themself. That being said, and if we accept the fact that rerunning the model is not possible for the present study, the authors have, in this revised version or their manuscript, addressed the comments they have received by modifying the text of the manuscript. They also have added, as a point of comparison to assess the performance of their model, new results from a non-convolutional neural network model as in Deneu et al (2021).

As a consequence, in this revised version, the manuscript is much clearer (e.g. regarding the classification task and species co-occurrences, regarding environmental variables) and the discussions is much more relevant scientifically and interesting for the readers. For instance, the Discussion section on "Suggestions to further improve the modelling methods" should be particularly inspiring for the readers.

Hence, since most of the comments raised previously have been properly addressed, I recommend this preprint after minor revisions.

See details below.

Evaluation on how the authors addressed my comments in particular

Here I list the main concerns I had in the previous version of the manuscript and my evaluation on how the authors have addressed them in this revised version:

1) Main concerns on the description of species occurrences and environmental variables in the Method section => The Method section is much clearer now. I still have one comment: the authors state in their response that "We provide the source for each variable in table 2, which includes their unit and many more information", while the units are still missing. Please add them in table 2.

Done

2) Problems identified in the predictions: the way co-occurrences are handled is clearer now and the way the results are presented and discussed has also been improved.

3) Limited relevance of the "SST+2°C" scenario: This problematic part has been removed in the revised version.

4) Choices of 38 taxa: I totally understand that the authors do not which to re-run their models and I am fine with their answers.

5) Relatively superficial discussion: the scientific relevance of the discussion section has been significantly improved.

6) The conclusion section also lacks strong scientific background: the conclusion has also been improved and is now satisfactory.

7) Minor comments: they have been addressed, except the following that was not understood: "Line 105: essential to?" => line 123, replace "which is essential to reproducibility." by "which is essential for reproducibility."

Done

Miscellaneous comments

Minor comments

As usually done, the authors could acknowledge the reviewers and editor/recommender for their help in improving the scientific quality of the manuscript. Such acknowledgments would seem especially fair in particular given the thorough inputs provided by Jean-Olivier Irisson.

Done (lines 510-513)

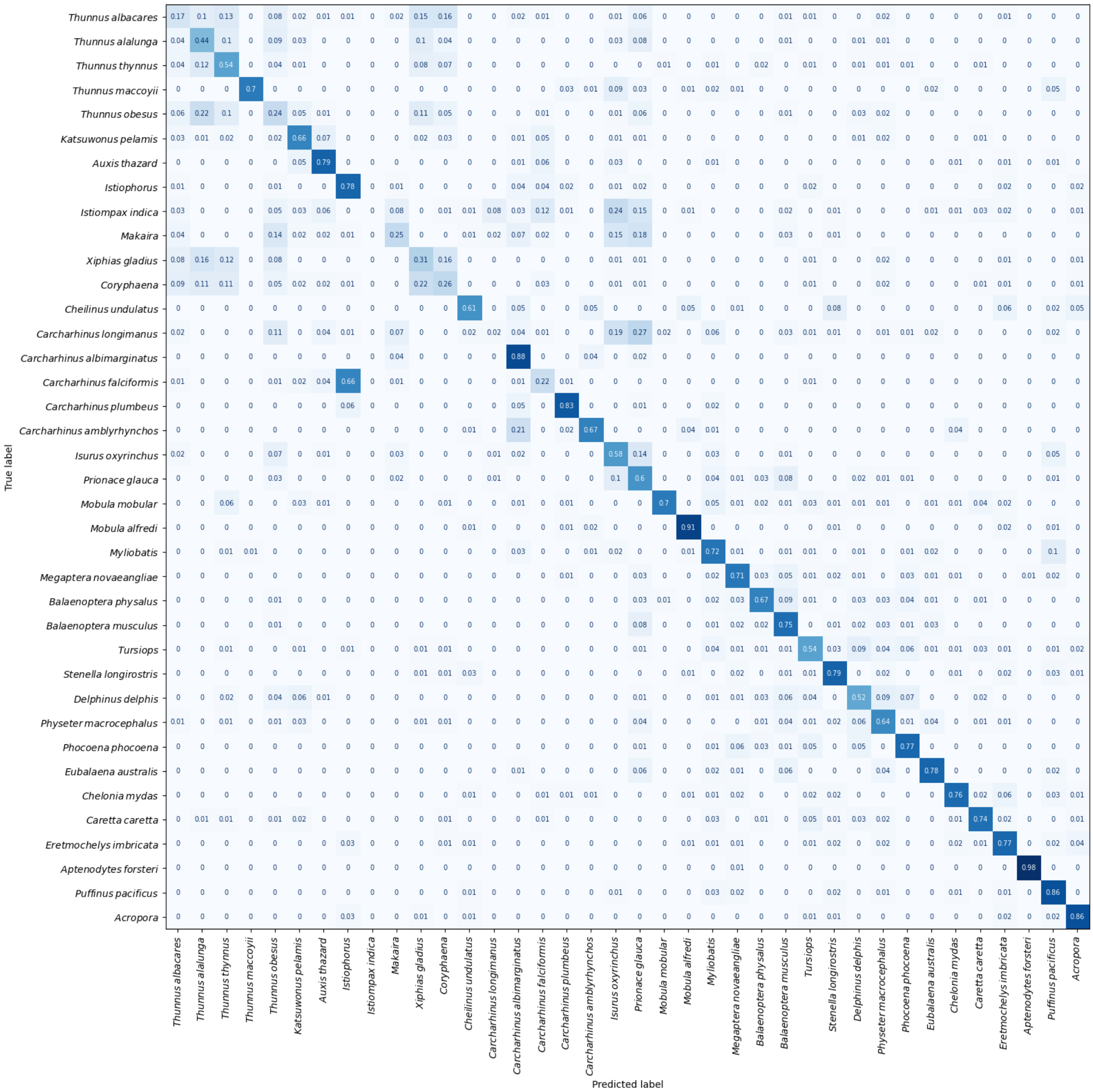
- line 171: "Because we propose a type of dynamic SDM, we cannot capture these long-term barriers, so we have to include them artificially." Alternatively, could you have considered using longitude and latitude? Or do you think it would have been a problem due to observation bias (as mentioned line 179)?

Indeed we cannot include geographical coordinates as it would lead to predictions gathering in sampled areas, therefore being limited by observation bias.

- Comparison between the CNN and punctual DNN: could you provide a rough estimate of the computing cost of each method? Is the confusion matrix similar? Are they some species that are better captured by the CNN, or is it due to a better performance of the CNN for all the species? Looking at Table 3, one could also argue that the increase in performance when using a CNN compared to a punctual DNN is not so high.

We did not optimize computing cost for the punctual DNN as we kept everything equal to the CNN (except input data) to allow a fair comparison of the results. So the computing costs are the same for both.
The DNN confusion matrix is very similar (included below), with all true positive rates lower or equal to the CNN (except *Mobula mobular* which is 0.01 higher).

Regarding the increase in performance, 6 percentage points is significant in the context of a 38 classes classification task. Users can decide whether it is worthwhile to them adding the extra complexity for these percentage points.



Punctual DNN confusion matrix

Suggestions for text editing:

- line 89: I suggest removing "as is usually the case with SDMs"

Done

- lines 96-97: Replace "Here we present an adaptation of their work that includes these adaptations to the specificity of the open ocean." by "Here we present an adaptation of the work of Botella et al. [22] and Deneu et al. [23] that includes these adaptations to the specificity of the open ocean."

Done.

- line 152: Add a reference to Table in 2 the sentence "the oceanographic landscape that we consider has limited precision due to environmental data resolution".

Done

- line 139: "Three of them contain two components: surface wind, geostrophic current and finite-size Lyapunov exponents (FSLEs)": do you mean zonal and meridional components? It seems later on (line 314) that it is rather polar coordinates ("strength" and "orientation"), but then in Tables 2 and 4 it is indicated that cartesian coordinates (u,v) have also been used. Please be more explicit line 139: "Three of them contain two values: both strength and orientation components (e.g., polar coordinates) for finite-size Lyapunov exponents (FSLEs), and both zonal and meridional components for surface wind and geostrophic current".

Done

- line 156: "115km" => "115 km"
Done

- line 164: "7km" => "7 km"
Done

- line 239: "100 km" => "100 km"
Done

- Figure 4 and Figure 8: If possible, the species names should be in italic.
Done

- Caption of Figure 5: "chosen because they deserve commenting" => "chosen to further discuss some interesting and contrasted patterns".
Done

- Figure 8: would it be possible to order the variables according to their order in Figure 7?
Done

- line 358: "A way to improve the final accuracy score would be to group species by traits": I am not so sure that species with similar traits are more truly observed more frequently together, since species also coexist in functionally diverse communities. Consider rephrasing, for instance "to group species by habitat preferences"?
Done

- line 403: "As previously discussed, this method would probably benefit from including a large number of taxa. In particular, planktonic species may prove valuable as they are less prone to sampling biases" => I would also argue that they are several databases of plankton species occurrence available for SDMs, as the ones developed at ETHZ, such as the 1704 species compiled by Righetti et al. (2020) or the 524 zooplankton and 336 phytoplankton species compiled by Benedetti et al. (2021) for SDMs:

We added this information (line 415).

D. Righetti, M. Vogt , N. E. Zimmermann, N. Gruber, 'PhytoBase: A global synthesis of open ocean phytoplankton occurrences', Earth System Science Data, 12, 907–933, https://doi.org/10.5194/essd-12-907-2020, 2020

Benedetti, F., Vogt, M., Elizondo, U.H. et al. Major restructuring of marine plankton assemblages under global warming. Nat Commun 12, 5226 (2021). https://doi.org/10.1038/s41467-021-25385-x

- line 430: Regarding the "spotted aspect of the map", I wonder if another pixel size for oceanic environmental variables could be more adapted for the 32 × 32 geographical pixels, compared to what has been done here inspired from terrestrial environments, e.g. for specifically targeting submesoscale features. Cf. for instance Lévy et al. (2018).

We chose this pixel size based on the available environmental data (cf. section 2.3.1 and Table 2). It would certainly be beneficial to use higher resolution data sources but they are not available for the whole world yet.

Lévy, M., Franks, P.J.S. & Smith, K.S. The role of submesoscale currents in structuring marine ecosystems. Nat Commun 9, 4758 (2018). https://doi.org/10.1038/s41467-018-07059-3

Review #3

Thank you to the authors for a very thorough response to my review and those of the other two reviewers. I felt that the discussion on both sides of the review was very helpful and has significantly improved the revised version of the manuscript. I have reviewed the responses to each of the reviews and the revised manuscript and am overall happy to recommend it be advanced (with minor/line edits) to acceptance from my perspective.

My largest issue (which was echoed by Reviewer 1) was related to the formulation of the prediction problem as a multi-class classification problem. I appreciate the authors' response to this point in their responses to both reviews and the addition of the new text highlighting this difference between their approach and a more traditional approach. I understand the authors' argument that this perhaps generalizes to predicting the relative detection rates of each species though I think the assumptions needed to get there are unrealistic in the general case. Regardless, I still believe this family of frameworks are on the rise and am happy to see this paper contribute to the general discussion of the technique as a whole, given that the authors have added a good amount of clarity around their methodology here.

Reviewer 1 makes several arguments about the focus/scope of the method and discussion, which I agree with and think the authors have done a satisfactory job resolving. I believe the revised draft of the manuscript is more concise and clear as a result and at no significant cost to its potential impact.

Reviewer 1 also makes numerous suggestions which could improve the analyses and which the authors have generally decided to incorporate into the discussion to setup future research lines for follow on work. I am of the opinion that in general these potential improvements are appropriate for future work and generally should not preclude the publication of this work and appreciate the authors' effort to transcribe these suggestions into the manuscript.

Given that the authors have indicated their aversion to significant revisions to their methods or results, I'd like to say that I am personally satisfied with the experiments that they've performed and am not overly concerned that their resulting discussion is misleading or that a reader might be misled about the approach as a result of the experimental setup the authors' have presented.

Line items
L91: optic → optical
Done

L93: Chlorophyll → chlorophyll
Done

L144: rephrase → "were encoded as layers with equal dimensions to the other variables."
Done

L190: typo: explicitely
Done

L273: cut
This sentence was added to answer a comment from the first round of reviews

Fig5 caption: cut "chosen because they deserve commenting."
The sentence was modified following a comment by reviewer #2

L329: coherent → consistent
Done

L351-355: I’d cut this attempt to justify any poor accuracy of a dynamic model via movement or individual stochasticity
We would like to keep this paragraph to remind the reader of the practical limits of dynamic distribution modelling.

L426: “one if the” → “one is the”
Done

Once again, thank you to the authors', reviewers 1 and 2, and the editors for the interesting paper and discussion. I again appreciate the opportunity to review this very nicely written manuscript.