[ArticleID #699 – Version 1] 'Using informative priors to account for identifiability issues in occupancy models with identification errors'

RESPONSES TO RECOMMANDER AND REVIEWERS COMMENTS

Response to Recommender

Reply: We are grateful to the Recommender for considering our manuscript and for having found referees. We have duly considered all comments and suggestions from both reviewers and we provide below detailed responses to each comment received. In the manuscript, we have highlighted all relevant changes in yellow. We believe the revisions significantly improved the contents and readability of our paper.

> Abstract needs revising (also see reviewer comments): although the middle part provides a nice summary of the problem at hand, it takes too much space and the abstract is slightly imbalanced as the simulation study is only mentioned on the fly and no results and lessons learned (recommendations) are provided.

Reply: We have shortened the middle part and added a sentence to sum up the contributions of our approach**.**

> Introduction:
>
> - the streamlining in the introduction could be improved to make very clear that occupancy-detection models have been around for some time but the problem of false positive (identification) problem is rarely addressed although becoming more and more relevant with the novel types of monitoring data. All this information is provided in the introduction but currently L7-25 read a bit like a history of occupancy models instead of clearly stating that this is the status quo but new data provide new challenges.

Reply: We have revised the introduction to first present occupancy models and new monitoring methods **(L 6-46)**. We have also added a paragraph to specifically address false-positive errors **(L 47-55)**, following the distinction between errors arising from data collection through novel monitoring techniques and their subsequent treatment.

> - L41-51 feel a bit out of place and I would recommend some restructuring with the previous paragraph to provide a clear story arc ranging from novel data peculiarities (image classification, acoustics, eDNA, etc) to mis-identification, which then can lead smoothly to models accounting for false positives (L54 onwards).

Reply: Following this suggestion, we restructured the introduction in a way to explain clearly and concisely the occupancy model for non-familiar readers **(L 6-13)**. Then we highlight data from new biomonitoring technologies **(L 14-46)**, the related misidentifications **(L 47-55)** and the adapted statistical tools **(L 56-67)**.

- in L56-72 the authors discuss more conventional solutions for addressing false positive like "reference sites". However, this part does not reflect on peculiarities of novel monitoring/sensor data that often rely on AI or bioinformatics approaches for classification. As the motivation for this study hinges so much on novel sensor data, it would be helpful to also mention more explicitly here why these data in particular may have increased levels of false positives (see e.g. https://doi.org/10.1016/j.tree.2023.09.017).

Reply: We appreciate this important suggestion and the example provided that we missed. We have added some sentences in the introduction part about false positive misidentifications **(L 50-55)** (we now cite Hartig and colleagues' paper, thanks for the reference) and later before introducing our extended model **(L 108-114)**.

- L79-83: as the study also uses a simulation approach to test the different approaches, this objective should be explicitly stated here.

Reply: We have added a sentence at the end of the introduction to state this **(L 86)**.

Structure/Presentation:

As voiced by #1, the structure could be improved and clearer signposts could help directing readers. I want to echo this and encourage the authors to more clearly separate different types of sections, e.g. model formulation (or concept) from simulation study. The simulation study could also follow a more classical structure of first explaining the methods and then the results. However, with clearer signposts, the more narrative structure could also work.

Reply: We agree that our narrative structure is a bit unconventional and was difficult to follow. Based on your recommendations, we have revised the section titles and added subsections to improve clarity. We also hope that the improvements in the writing will help guide the reader through the structure more effectively.

- L135-225: these sections use very short paragraphs containing only 1-2 sentences. As result, the text may appear incoherent and like a list of unconnected thoughts. Here, the text should be revised and coherence be improved, also better balancing paragraph size.

Reply: As pointed out by both reviewers, we agree that the quality of the writing had to be improved. In addition to revising the text and the transitions between paragraphs, we have renamed this section “Simulation study”. We have rearranged it into several subsections to sequentially deal with the classical frequentist approach, the identifiability issue and approaches to address it. We have also better balanced paragraph size.

Figures:

Not all Figure captions are stand-alone, meaning they do not provide enough information to understand the content of the figure without consulting the manuscript text. For example, Fig. 3 caption does not explain the notations (e.g. w_A and p_hat), and does not explain "with/without constraint" (similar issues with Figs. 1 and 4, and appendix figures). By contrast, Fig. 2 caption is much more complete and stand-alone. Please carefully check and revise all figure captions.

Reply: We agree the captions had to be improved, so we modified them all (except that of Fig. 2).

Response to Reviewers #1 and #2

Reply: We are grateful to both reviewers for the thorough analysis of our manuscript and the insightful comments. Overall, we believe the revision has greatly improved the contents and readability of the paper, and we hope that all concerns of the reviewers have been fully addressed.

Response to Reviewer #1

Writing
Given the need for clarity in such a methods-focused paper, the writing in this article could be made more direct and concise – there are areas where meaning is difficult to interpret due to the phrasing. More specifically, many sentences are overly long with many commas and could easily be separated into independent clauses.

Reply: We agree that the writing quality affected readability, so we have thoroughly revised and reshaped the entire text.

Abstract
I think that this abstract would be challenging to interpret for people who are not very familiar with the occupancy model and particularly its false-positive extensions. These models and their limitations in their standard form need to be more generally introduced. It may also be slightly too long, with some content more appropriate for the discussion.

Reply: We thought that our paper would mostly interest people already familiar with occupancy models, that is why we did not focus on its basic formulation in the abstract (in the style of Guillera-Arroita et al. 2017 and Chambert et al. 2018). However, we agree with you, and we have tried to introduce the topic in a more general manner in the *introduction* **(L 6-13)** and in the *Model Description* **(L 93-103)** section. Also, as you and the recommender suggested, we shortened the abstract to focus on our contribution.

Article structure
The layout of the manuscript, including headings, is unconventional and make it challenging to interpret. It is difficult to tell where the introduction and model formulation end and where the simulation studies begin, and within the sections 'Classical estimation from the identification layer (L135)' and 'Using an informative prior to address identifiability issues (L176)' the methods, results, and discussion components are heavily intertwined. These should be more explicitly separated; one example of where this occurs is L140-152, where elements of methods, results, and discussion occur in a confusing order.

Reply: We acknowledge that our structure is unconventional. Following your recommendation, we have reordered several sentences and added transition sentences to make the articulation of sections and ideas more consistent. Please see also our reply to the recommender on this topic.

Figure layout could also be improved (e.g. Figure 3 should be moved to the prior section), and references to figures in the supplement (as in L169) should more explicitly note when figures are supplementary. The caption for Figure 1 should also be expanded to include definitions of the included parameters, as it currently cannot be independently interpreted.

Reply: We agree, and we have improved captions for all figures (except that of Fig. 2).

Occupancy models and passive sensors

At times the paper seems to overstate (likely inadvertently) the connection between occupancy models and passive sensor data. This is most prominent in the abstract and introduction (L1-25), but occurs elsewhere too (e.g. L97). More clearly stating a) what occupancy models are, b) how passive sensor data differs from other presence/absence data types, and c) what issues those differences produce could help to clarify this.

Reply: It is true that the connection between occupancy models and passive sensor data is central in the article. This is because we introduce, as early as in the abstract, the contrast between detection and identification, which is particularly relevant to passive sensor data. We do not intend to consider passive sensor or eDNA sampling as prerequisites for modelling occupancy, but we believe that they are far more likely to introduce identification errors than other types of sampling. We have followed your recommendation, and as suggested by the recommender too, we have added a paragraph on it in the introduction **(L 46-55)**.

Model formulation

The parameterisation of the occupancy model is described nicely, but it should be more clearly delineated in the text where the model diverges from the standard occupancy model (c. L94). Describing the detection and identification processes in more general terms rather than with respect to passive sensors may also be appropriate, as there is little reason why this model could not also be applied to standard field surveys performed by humans.

Reply: Thanks for your suggestion. We have explicitly separated in two subsections the original model from the extended version we are proposing. Also, we have avoided to talk about "records" or "classification" during the model formulation **(L 105-132)**.

Bayesian models

The methods for the Bayesian simulation study require more detail in the main text. Most importantly the priors used for occupancy probability, detection probability, and false identification probability must be included.

One of the biggest questions many readers will ask is how sensitive this model is to bias induced when the informative priors used are *not* appropriate for the data – this is mentioned in L76-77 but is not explored in the simulations nor further commented on in the discussion. This should be expanded upon, at least with respect to limitations in the discussion.

Reply: We have added a general comment about using an inappropriate prior, as we think this is a general limitation when using informative priors in Bayesian studies **(L 345)**. Following your recommendation (see also a comment by reviewer#2), we have added in the appendix a sensitivity analysis regarding the possible values that the positive identification parameter can take in this model **(L 387-393)**.

Conclusion

The conclusion does not feel specific to the model defined in this article. It could be more explicitly stated how this manuscript contributes to the management of the identifiability issues commonplace with passive sensor.

Reply: We have added in the discussion paragraphs about the limitations of our approach **(L 328-334)**.

Detailed comments

Reply: We have addressed all detailed comments with no exception. Thanks a lot.

Abstract: *"the naïve occupancy model does not account for false detection"*
Specify that this model does not account for false *positive* detection to increase clarity, as false negatives are accounted for.

Reply: We have specified that alternative models attempt to account for false positive detections.

Abstract: *"Overall, what is at stake is enhancing statistical methods together with sampling noninvasive technologies, in a way to provide ecological outcomes suitable for conservation decision-making."*
This sentence is a bit strangely phrased; 'what is at stake' could be replaced with "the objective of this article is to …"

Reply: Thank you for the suggestion, we have rephrased the whole last sentence.

Caption for Equation 1: *"… formulation of the occupancy model (Royle and Kéry, 2007) …"*
It should be noted that the citation for *Royle and Kéry 2007* is for the *multiseason* implementation of the occupancy model, although this part of the formulation is the same for the single and multiseason versions.

Reply: We have deleted the reference in this sentence.

L149 *"It has been showed …"* it has been shown

Reply: Thank you for this correction.

L193-196: "*We study 3 types of priors for parameter wA … We introduce 2 different prior distributions for the probability to correctly identify the species …"*
These two sentences are unclear and seem repetitive. The first says three types of priors are used, the second that two prior distributions are used. This could be simplified to reduce ambiguity.

Reply: We acknowledge this section was unclear so we have entirely modified it.

L239 *" … and above all they are not specific, …"*
This phrasing is somewhat unclear; maybe "are not species-specific"?

Reply: Thank you for the suggestion, we have reworded this paragraph **(L-288)**

L253: *"…by reducing data processing time, potential identification errors are introduced …"*
Further elaboration required on how this may occur

Reply: We have clarified this sentence, with "automation without human validation" **(L-303)**

L260: *"… we need feedback on the performance of the identification process …"*
'Feedback' doesn't quite fit in the context – 'information' or something else may be more appropriate.

Reply: Thank you for this comment, we used the word of "knowledge" instead **(L-328)**.

L266 *"involving inputs on the detection process in the form of a prior"*
There is somewhat more consistency needed on separating the 'detection' and 'identification' processes – in this article, only priors on the *identification* parameter appears to be included, not detection.

Reply: Thank you for pointing out this error, we have corrected it **(L-343)**.

Response to Reviewer #2

In this paper, Monchy and co-authors present a variation of the widely used site-occupancy model that attempts to deal with the challenge of false-negative and false-positive errors. The motivation is to provide a method that can be used to analyze large volumes of occupancy-detection data provided by newer sampling methods such as camera trapping and eDNA collection. The model adds a third level to the hierarchical occupancy model, identification. Species can be mis-identified in either direction. A species that is present on site and detected (e.g., in a photo) can fail to be properly identified and therefore erroneously treated as absent. The converse is that a species that is absent from a site can fail to be detected (e.g., no photos are taken of the target species) but still identified as present if the sample is mis-interpreted. The authors use data simulation to compare several approaches for analyzing data with the three-level occupancy model and quantify biases induced by species mis-identification. First, they use maximum likelihood to try to estimate ψ, p, wA, and wB for N=30 sites and either J=12 or J=36 visits per site. Increasing the number of visits improved accuracy for estimates of ψ, but estimates of p and wA were biased when fitting the three-level model with identification via maximum likelihood. Next, the authors simulated data with varying values for wA and wB, and estimated occupancy parameters using maximum likelihood and either 1) no constraint on wA and wB or 2) a constraint such that wA > 1-wB. This constraint did not completely resolve the identifiability issue. Finally, the authors fit Bayesian three-level occupancy models with either uninformative or highly informative priors for wA. When priors were uninformative, estimates of p and wA were biased, whereas models fit with the two informative priors both produced much more accurate estimates for these parameters.

I enjoyed reading the manuscript and I think it seeks to address an important issue that does not receive enough attention in occupancy studies. Particularly with the increase in massive datasets from these newer sampling methods, false positives (and new types of false negatives) will take on increasing relevance for interpreting occupancy data. I also applaud the authors for exploring using informative priors in a Bayesian model, which I think should be more common in ecological studies. My main comments are: 1) I would like to see more explanation of the prior elicitation process and the data used to obtain these priors (ideally it would be shown in the accompanying R code as well). 2) I think the authors should address if there is any way the data used to estimate the correct classification rate for photos could be directly incorporated into the Bayesian occupancy model as another observation process. 3) I wonder if the apparent benefit from the informative priors is dependent on the true value of wA being very high (~ 0.9), and what this says about the value of including the identification process. The manuscript lays out a good case for why it is very important to model the identification process for some of these semi-automated sampling methods, but then ends by stating that sensitivity is often very high (at least for classifying photos) and that priors that constrain wA to values near 1 improve estimation. If wA is nearly 1, does that somewhat refute the need for modeling the identification process? I explain each of these comments in more depth below. I feel addressing these issues would improve the manuscript and help make this manuscript a more valuable contribution to the literature.

Major comments

-Lines 117–118 – If the assumption of the traditional occupancy model is that the false positive rate (1-wB) is zero, should wB = 1 here, so that the false positive rate (1 – wB) = 0? Looking at the comment in line 17 of the code, w_b is defined as the "correct non-identification prob" so I think wB should be 1 if non-identification is assumed to be impossible in the traditional model. I could be confused but I think Figure 1 aligns with my interpretation.

Reply: Totally true! Thank you for pointing out this error.

-The manuscript contrasts the new method with other approaches to dealing with false positives that require a secondary data source with a known true detection/identification rate (Miller et al., 2011). But could you consider the results on the accuracy of an image classifier (e.g., lines 197–203) a secondary data source as well?

Reply: We consider the accuracy as a required information but the whole dataset enabling to get this information is not necessary (and in general not available to us), that is the reason why we do not consider it as a secondary source of data. We think the classifier is for example an opensource tool already trained and that its performance might be available to ensure its transferability on an unlabelled dataset. We have added elements about that in the *"Using an informative prior"* subsection **(L 205-229)** and in the discussion **(L 328-334)**.

-The informative priors used to address identifiability in a Bayesian three-level occupancy model are based on data relating to the sensitivity of the identification process. Given the use of a hierarchical Bayesian model, and the existence of at least some data to estimate sensitivity, could you incorporate those data into an observation process sub-model within the larger occupancy model? In other words, rather than using the metrics that are the output of a classification algorithm (from a neural network) could one take the validated data on classification, feed it into a Bayesian model of the identification process, and use that to estimate wA directly in the model? Even if it were a subset of the data that were validated (e.g., a neural network identifies all photos, humans confirm a subset of the data), that could be informative when plugged into a sub-model for identification. This is alluded to on lines 181–182: "For example, one may simply use a dataset that has been annotated during a previous manipulation" Why not use that dataset directly in the model rather than eliciting a prior from those data, then putting that prior into the model? In a Bayesian hierarchical model, I can imagine how this just becomes another sub-model representing the identification process.

Reply: In our situation, the data is processed automatically, which may introduce unknown errors. Therefore, we aim to account for misidentifications. While neural networks are tested to ensure they provide reliable classifications, a verified dataset is needed for these performance tests. However, in this context, we do not have access to manually labeled, error-free data required to estimate sensitivity, so we are unable to incorporate it into the model.
We have clarified this point **(L 208-214)**

-Lines 197–218 – I have two comments on the use of informative priors to improve identifiability:

1) I feel the description for how these priors were chosen is lacking. Sources were cited for prior elicitation, but at least for eDNA these sources are general references on prior elicitation, not specific to eDNA data as far as I can tell. Reference is made to using a matching method to choose a beta distribution "with the appropriate values arisen from the survey results." What survey results does this refer to? Could these results and the prior elicitation be included in the R code that accompanies the paper? I appreciate the way the R code for the simulations was shared, but this part of the analysis is quite opaque. For the camera trap prior elicitation, the paper cites Gimenez et al. (2022) in an earlier section to explain how parameter values were chosen for the simulations. Could you devote a little more text here to explaining how the previously published data on photo classification was used to select the Uniform(0.8,1) prior?

Reply: We agree that the argument for the prior elicitation was light, especially for eDNA sampling. We now specify how this prior needs to be chosen according to the available knowledge, which is quite different between eDNA data and sensor data. By looking for references where the sensitivity was reported without using additional unambiguous surveys, we found that the sensitivity of eDNA identification was not transferable from one study to another, unlike the classification of sensor data. The only option is to conduct positive controls to get this value of sensitivity in eDNA-based studies, however this is often lacking. As a consequence, and to clarify this section, we have decided to focus on sensor data classification for prior elicitation. We only discuss the possibility to use this informative prior approach in eDNA-based studies providing that knowledge is available. And broadly, we discuss the importance for the providers of classifiers to make the performance metrics accessible to the community **(L 332-334)**.
To improve clarity in the elicitation process, we have associated each level of available knowledge with a specific type of prior and explained the context for its use **(L 230-250)**. Additionally, we have enhanced the R code to include numerical elicitation.

2) I would like to see a broader sensitivity analysis with priors that are not either completely uninformative or highly informative. Both informative priors have very high prior probability that sensitivity is very high (near 1), with a mean around 0.9 (which was the true value of wA used to simulate data). Not surprisingly, the models that use these priors tend to estimate posterior distributions for wA that are also centered near the true value of that parameter. What would the results look like if the true value for wA were much lower? What would the results look like if one used a prior that is somewhere between Uniform(0,1) and Uniform(0.8,1)? A stated problem with the original occupancy model is that it assumes *a priori* that wA = 1. If the conclusion of this study is that wA is usually quite close to 1 (e.g., for image classification of Lynx in Gimenez et al. 2022, recall was 0.95), does that diminish the need to account for identification somewhat? If all a researcher cares about is ψ, are the estimates of ψ in Figure 4A close enough? This point is mentioned on lines 261–263, but I think it deserves a little more discussion.

Reply: We appreciate this comment which raises several important points:

1. The different levels of informativeness

To address this, we have added one more prior to our simulations, so that we have now 3 alternatives to the default prior.

2. The true value of w_A

We agree that presenting a new hierarchical model requires to explain the range of plausible values for parameters. We have added in the appendix a sensitivity analysis with more values of w_A for data simulations.

3. The consequences to use a wrong informative prior

We agree, and this is a general topic in Bayesian statistics, not specific to our model; we have added a sentence to discuss this **(L 345-347)**.

4. The effect of the prior choice on the occupancy parameter

The need to account for identification remains even if w_A is close to 1. Indeed, false-positives misidentifications may abruptly overestimate occupancy probability (this result is already known and we relate to it the *introduction* **(L 58-59)** and in the *"classical estimation"* subsection **(L 160-163)**.

Finally, it is true that the estimate of ψ does not really change with the choice of the prior, so we qualified a bit more this point in the discussion **(L 335-340)**.

Minor comments

-The text could use a thorough editing pass by a native English speaker. I could understand the meaning of the text, but it is awkward in places such that I had to re-read statements to ensure I grasped what was being communicated. A minor example, is the frequent use of "allow to" or "allows to" (lines, 12, 22, 73) or "enable to" (second line on page 4) which could be replaced with simply "allows" or "enable the"

Reply: As pointed out by the recommender and reviewer #1, we agree that the quality of the writing was making it difficult to read the manuscript smoothly**.** The entire text has been reread and reshaped and the suggestions of language have been followed.

-Line 43 – I think this statement could be revised to more broadly describe eDNA methods, because not all eDNA studies use PCR (e.g., see the use of SHERLOCK methods by Nagarajan et al., 2024 here: https://onlinelibrary.wiley.com/doi/full/10.1002/edn3.506).

Reply: We agree that many genetic tools exist in eDNA studies, and we have broadened our statements about it in the *introduction* **(L 33-41)**.

-Line 101 – Change "deduct" to "deduce"

Reply: Changed accordingly

-Lines 146–148 – Here it states that the original occupancy model without the additional identification layer overestimates ψ. I think it is pretty self-evident, if all sites have at least one identification then the original model (which ignores false positives) would estimate ψ = 1. I would still like to see a comparison between the original occupancy model and the occupancy model with the identification layer. Just one additional sentence with the estimates of ψ and *p* from the original model would be helpful.

Reply: Following your suggestion, we have added a sentence to quantify how the ψ estimate is overestimated with the original model**.**

-Lines 169–170 – Figure 7 is referenced before Figure 3 or Figure 4. Likewise Figure 5 is cited before Figure 3 or Figure 4. I recommend re-ordering figure numbers to match the order they appear in the paper.

Reply: We appreciate this comment and have changed the reference numbers of the figures in the appendix.

-Line 247 – Change "requires to develop" to "requires developing"

Reply: Changed accordingly

-Line 272 – Change "allow to study species" to "enable studying species"

Reply: Changed accordingly

-Figure 4 – I wonder what the posterior distribution is for the product of *p* and wA? In panel A, when the Uniform(0,1) prior is used, *p* is overestimated and wA is underestimated. If you take their product, do the two errors effectively cancel one another out?

Reply: It is true that the biases seem to cancel one another out, but not entirely. The product of the estimates is still biased depending on the value of w_A.

-Figure 7 – Please explain the vertical red lines in the caption.

Reply: The caption has been detailed with the red line description.

R code

-Lines 87–88 -> The line creating res_simu12 is commented out and the object created is res_simu36 instead. I think this section is confusing, because some parts reference J=12 (line 113) and others reference J=36 (lines 105 and 110). The comment on line 85 says to run once for J=12 and once for J=36. But it is not immediately clear that this requires setting J=36 and running everything from lines 47–75 again. I know this is more concise, but I think it would be clearer for those reproducing the analysis if there was a second block where J is set equal to 36 and some code was duplicated (e.g., lines 47–75) to rerun the simulation and ML estimation with this larger simulated data set.

Reply: We agree that copying the block of code is more convenient than rerunning it with different settings. Also, we have split the code into 3 files in order to associate each file to a subsection of the *"Simulation study"* part.