- **Title and abstract**
    - Does the title clearly reflect the content of the article? [X] Yes, [ ] No (please explain), [ ] I don't know
    - Does the abstract present the main findings of the study? [X] Yes, [ ] No (please explain), [ ] I don't know
- **Introduction**
    - Are the research questions/hypotheses/predictions clearly presented? [X] Yes, [ ] No (please explain), [ ] I don't know
    - Does the introduction build on relevant research in the field? [X] Yes, [ ] No (please explain), [ ] I don't know
- **Materials and methods**
    - Are the methods and analyses sufficiently detailed to allow replication by other researchers? [ ] Yes, [X] No (please explain), [ ] I don't know
        - See comments below regarding the description of the prior elicitation.
    - Are the methods and statistical analyses appropriate and well described? [X] Yes, [ ] No (please explain), [ ] I don't know
- **Results**
    - In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)? [ ] Yes, [ ] No (please explain), [X] I don't know
        - Not relevant
    - Are the results described and interpreted correctly? [X] Yes, [ ] No (please explain), [ ] I don't know
- **Discussion**
    - Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? [ ] Yes, [X] No (please explain), [ ] I don't know
        - I think the implications of using such highly informative priors are not completely explored, see comments below about prior elicitation and possible comparison of models will less informative priors.
    - Are the conclusions adequately supported by the results (without overstating the implications of the findings)? [X] Yes, [ ] No (please explain), [ ] I don't know

**Review of Monchy et al. (2024)**

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 $\psi$, $p$, $w_A$, and $w_B$ for N=30 sites and either J=12 or J=36 visits per site. Increasing the number of visits improved accuracy for estimates of $\psi$, but estimates of $p$ and $w_A$ were biased when fitting the three-level model with identification via maximum likelihood. Next, the authors simulated data with varying values for $w_A$ and $w_B$, and estimated occupancy parameters using maximum likelihood and either 1) no constraint on $w_A$ and $w_B$ or 2) a constraint such that $w_A > 1-w_B$. 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 $w_A$. When priors were uninformative, estimates of $p$ and $w_A$ 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 $w_A$ 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 $w_A$ to values near 1 improve estimation. If $w_A$ 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 ($1-w_B$) is zero, should $w_B = 1$ here, so that the false positive rate ($1 - w_B$) = 0? Looking at the comment in line 17 of the code, w_b is defined as the "correct non-identification prob" so I think $w_B$ 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.

-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?

-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 $w_A$ 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.

-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?

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 $w_A$ used to simulate data). Not surprisingly, the models that use these priors tend to estimate posterior distributions for $w_A$ 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 $w_A = 1$. If the conclusion of this study is that $w_A$ 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 $\psi$, are the estimates of $\psi$ in Figure 4A close enough? This point is mentioned on lines 261–263, but I think it deserves a little more discussion.

*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”

-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).

-Line 101 – Change “deduct” to “deduce”

-Lines 146–148 – Here it states that the original occupancy model without the additional identification layer overestimates $\psi$. 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 $\psi = 1$. 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 $\psi$ and *p* from the original model would be helpful.

-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.

-Line 247 – Change “requires to develop” to “requires developing”

-Line 272 – Change “allow to study species” to “enable studying species”

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

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

*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.

**References**

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L., Weir, L.A., 2011. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. Ecology 92, 1422–1428. https://doi.org/10.1890/10-1396.1

Nagarajan, R.P., Sanders, L., Kolm, N., Perez, A., Senegal, T., Mahardja, B., Baerwald, M.R., Schreier, A.D., 2024. CRISPR-based environmental DNA detection for a rare endangered estuarine species. Environ. DNA 6, e506. https://doi.org/10.1002/edn3.506