

Summary

The manuscript ‘Using informative priors to account for identifiability issues in occupancy models with identification errors’ introduces a new parameterisation of the single-season occupancy model with parameters for occupancy, detection probability, true-identification probability, and false-identification probability. The model is applied to simulated data in the maximum likelihood and Bayesian frameworks, with informative priors for true-identification rate used in the latter. The approach introduced here differs from other occupancy models extended to account for false positive detections in that it does not require additional datasets or a source of true positive detections in the data. As the authors note, this is particularly applicable to studies using camera traps, acoustic monitors, or eDNA sampling – all increasingly popular data sources for occupancy models.

General comments

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.

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.

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.

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.

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.

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.

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.

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.

Detailed comments

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.

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 ...”

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.

L149 "*It has been showed ...* "

"it has been shown"

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.

L239 "*... and above all they are not specific, ...* "

This phrasing is somewhat unclear; maybe "are not species-specific"?

L253: "*... by reducing data processing time, potential identification errors are introduced ...* "

Further elaboration required on how this may occur

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.

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.

PCI Questions:

- Does the title clearly reflect the content of the article? **Yes**
- Does the abstract present the main findings of the study? **Yes**
- Are the research questions/hypotheses/predictions clearly presented? **Yes**
- Does the introduction build on relevant research in the field? **Yes**
- Are the methods and analyses sufficiently detailed to allow replication by other researchers? **No - the supplementary R script provided is sufficient, but the main text requires further details on all priors used for the Bayesian implementation.**
- Are the methods and statistical analyses appropriate and well described? **Yes**
- Are the results described and interpreted correctly? **Yes**

- Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument? **No - further discussion is needed on the potential implications of poorly-defined informative priors.**
- Are the conclusions adequately supported by the results (without overstating the implications of the findings)? **Yes**