February 28 2025

Dear Prof. Steinbauer,

We are pleased to submit a revised version of our manuscript *Twenty years of dynamic occupancy models: A review of applications and look to the future.* We greatly appreciate the constructive feedback provided by the two reviewers and thank them for their recommendations, particularly regarding the structure of our manuscript and aspects of our discussion. To create a more cohesive and targeted narrative it was necessary for us to make considerable changes to the structure and text of the article. These changes include a return to more conventional results and discussion sections and the reorganising of the discussion to centre around clear recommendations and research priorities. We believe that these changes have improved the clarity of our review and thank the reviewers for their suggestions. Our responses to both your and the reviewers' comments are outlined below, with the original comments in italics, our responses in standard font, and line numbers indicated with the "L" prefix.

Thank you for the opportunity to re-submit this manuscript.

Best regards,

Saoirse Kelleher (on behalf of co-authors)

EDITOR'S COMMENTS

Dear authors,
The manuscript has been reviewed by two experts who agree on the relevance and quality of your work but also outline why it is not suitable for publication in its current form. Both reviewers point out that the manuscript draws vague conclusions and fails to offer specific directions to readers. Its relevance would significantly benefit from deriving more precise recommendations for the future development and application of the models. The suggestion to structure the manuscript around a few well-defined objectives seems to be a promising way to provide clearer insights for readers.
I am looking forward to a revised version of your manuscript.

- We appreciate the reviewers' comments on the recommendations made in the original manuscript. To improve the focus of our review, we have made major changes to the overall structure of the article to recentre it around more clearly delineated objectives in L71-82. These objectives are now reemphasised as subheadings in the amended methods and results sections.
- Our discussion is now structured around several areas which we believe represent particularly challenging parts of the modelling process for users of DOMs. In each of these 5 focal areas, we emphasise practical recommendations that are supported by the literature and highlight research priorities where uncertainties remain. These recommendations and priorities are re-emphasised with a summary in the new Table 2.

REVIEWER 1'S COMMENTS

GENERAL
The manuscript summarizes a sample of papers using dynamic occupancy models. The sample size is not huge (for understandable reasons), so only broad-stroke comments can be made. Despite this, the manuscript does raise some interesting issues.

- Thank you for taking the time to provide useful comments and suggestions.

The main problem I have is that it raises issues, but doesn't really grapple with them: the discussion doesn't get much further than Something Must Be Done!

- We agree that the recommendations in the original manuscript were often too vague to be practicable. In the amended version, we have written a new discussion section organised around 5 key topics in fitting DOMs. In each of these we make recommendations and indicate research priorities that we believe are more practical.

For example, there is a mention that few studies use a robust design, but what is the consequence of this? Is it a problem for the studies? I can see that there could be cases when it is a problem, but also cases when it is not. The authors have looked at almost 100 papers, so they should be able to assess how much of an issue it is. Similarly, I was disappointed that there was no discussion of the biological appropriateness of DOMs. They are clearly appropriate for a classical meta-population, but what about (for example) occupancy models on whales, living in what is essentially a continuous habitat? It doesn't seem like the right approach, but of course it might not make a lot of difference. The closure assumption is a similar issue: sometimes it is reasonable, sometimes it is not, but when should one worry about it? Even a couple of sentences laying out the issue, and saying where to go for more help would be useful

- We agree that discussions around the robust design, closure and continuous habitat merited considerably more detail and have focused on trying to unpack these issues for readers, including providing guidance on when and how they may pose problems. The first section of the new discussion (Defining occupancy and detection; L354-434) covers these topics. We highlight key parts of this new text below.
- We now highlight when violations of the closure assumption are likely to result in biased estimates and point to texts that provide guidance on how to identify and avoid these violations.
    - *L381-388: "Earlier occupancy model research was often focused on conventional definitions of occupancy which require compliance with the closure assumption. This work emphasised how violations of this assumption can bias estimates of model parameters, leading to overestimates of occupancy when sites are not consistently occupied within primary occasions due to non-random mortality or movement of individuals (Otto et al., 2013; Valente et al., 2017). These authors provide guidance for study designs to ensure closure, as well as tests and analysis-based solutions for cases where closure is violated (Kendall et al., 2013; MacKenzie et al., 2017; Rota et al., 2009)."*
- Aided by a new figure (Fig 7), we outline how study design (e.g. continuous V discrete habitat patches, size of sites relative to home ranges) influences violations of the closure assumption and interpretation of model outputs (L362-371), as well as considerations for model building (L402-415). We frame this more of an issue of interpretation meriting caution, rather than a concern regarding biological appropriateness. We use examples and our review results to make this discussion more concrete e.g. we illustrate situations when model outputs likely represent use (i.e. the probability that a species is present at a site *at some point* during a primary occasion) rather than more conventional interpretations of occupancy.
    - *L394-401: "Many DOMs in our review would best fit this definition of occupancy: closure is often an unreliable assumption when modelling the mobile birds and mammals which dominate our sample, particularly when sites do not align with species home ranges. When data is collected with autonomous units like camera traps or acoustic monitors, small detection ranges make violation of closure very likely, necessitating a definition of occupancy which represents the probability that a species uses habitat that overlaps with the unit's detection range during the primary occasion (Wood & Peery, 2022)."*

- While we do not mention the robust design by name, we do comment on the questions around structuring data not collected under the robust design and highlight it as a research priority.
    - *L416-427* "An unusual challenge arises when users seek to fit to DOMs to detection/non-detection data that is continuously recorded – this may occur with camera traps or acoustic monitors which can run for long time periods, or with data from citizen science portals that lacks clearly delineated survey periods. In these situations, modellers must arbitrarily determine how to stratify detections into primary and secondary occasions (but note the availability of continuous detection models for secondary "occasions"; (Emmet et al., 2021; Guillera-Arroita et al., 2011; Pautrel et al., 2024). This stratification affects the temporal scale at which occupancy is measured, with corresponding changes to the definitions of occupancy and detection. The appropriate duration of these occasions will of course depend on

research objectives and the ecology of the target species (Chave, 2013), but guidelines for approaching this task in DOMs specifically would hold value for the increasing numbers of authors fitting models to this type of data."

The data are broken down into time versions, but there is little use of this. In art this might be a power issue: with 20 data points per period, there is not a huge amount of information once you start breaking it down.

- We agree that more could be done with this and have added a section on changing practices in DOMs over time (L321-341, Figure 6). Trends highlighted include numbers of sites, covariates consider, modelling techniques, and evaluation rates. As noted, our sample sizes are not large, and we explicitly acknowledge this in the text:
    - *L322-323: "… modest sample sizes within strata necessitate caution in interpretating these results*"
- These trends are not necessarily used to justify recommendations but rather are used to guide which areas we believe are the most important to focus our discussion. Due to these objectives, we believe that any spurious trends due to the sample size pose limited risks.

SPECIFIC

L101: Do you mean closed within seasons? i.e. no change in occupancy status in the sampling period?

- This typo has been fixed in Box 1:
    - "sites *are closed to changes in true occupancy state within each primary occasion"*

L150: 2020-2024 or 2020-2023? Also, when did you search Web of Science?

- The last strata was from 2020-2023; typo corrected. The date of the WoS search (July 26 and 29 2024) has been added in L94-95:
    - *"A pool of candidate articles was generated using two queries on Web of Science on July 26 and July 29, 2024"*

L151: Is this sample size large enough? Now you have gone through the papers, can you more quickly increase the sample size by focusing on relevant properties of the papers? I'm not saying this has to be done, but it might be fairly quick to pull out some of the data on all of the papers.

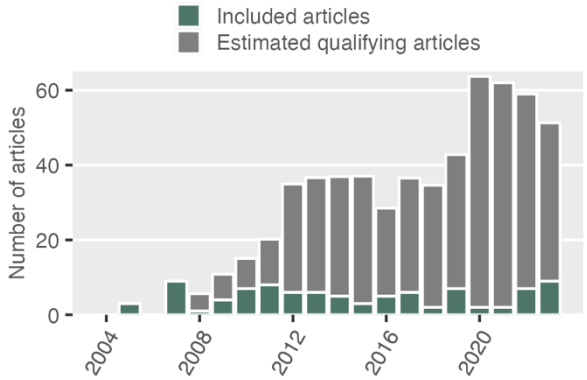
- While adding more papers would be possible, we believe that the key results from the current sample are sufficiently clear to support our findings and recommendations. As previously stated, these results largely inform the topics we identify as key challenges, and thus reducing uncertainty would not substantially change the manuscript.

L198: What happened to the other 8 papers (see L150-151)?

- The first strata (2004-2007) contained only 12 qualifying articles. This has been clarified in L165-166:
    - *"92 articles were included for this review, fewer than the 100 possible articles due to a deficit of qualifying papers in the first stratum (2004-2007)."*

Figure 3: What are the 4 bars in each segment? If they are years, it would be clearer to have a continuous (well, integer) X axis, rather than than 5 discrete classes.

- The bars in the previous version were years, we had labelled the axis as strata to clarify that the estimated articles were calculated using acceptance rates at that level. We have adjusted Figure 1 and clarified the caption accordingly:



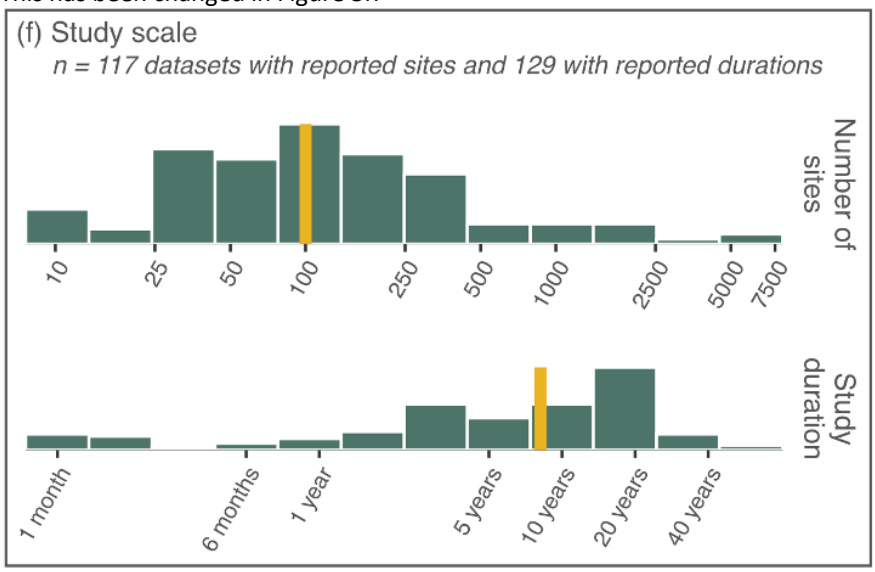
-
- *"…Projections were calculated using the number of unprocessed articles remaining in each year multiplied by the acceptance rate of the corresponding stratum"*

L216-7: What is the "primary occasion"?

- We have removed all references to 'season' and replaced them with 'primary occasion' to clarify that we are referring to the sampling window in each time step and to avoid confusion with ecological seasons.

Figure 4: "Site quantity" should be "Number of sites". Also, how does "used human observations" differ from "used citizen science data"? You seem to be implying that citizens are not human!

- This has been changed in Figure 3f:



-
- Citizen science was a subset of human observations. We now refer to human observations as 'direct observations', and clarify the relationship in L221-222:
  - *"79% of studies used direct observation data, 11% used live-trapping methods, and 14% used detections from camera traps. Within these categories, 10% of studies used citizen science data."*

L271: Should there be a new section heading here? You seem to have gone from Results to Discussion. I don't mind you doing this, but it would help if it was signposted.

- On reflection we see that the former structure was confusing, and have returned to a more conventional introduction/methods/results/discussion format.

L271-2: I think you should mention robust design frameworks before this. It's not clear what you mean here, and the following sentences leave the reader to work out what what you are writing about, but not why this is a problem.

- We have added a discussion of the data structure used for DOMs to the introductory Box 1. We no longer explicitly reference the robust design, instead discussing closure violations as previously mentioned.

Table 1: This is big and horrible. Some aspects are not clear, enough. if 14% of papers used hydrological covariates for any parameter, how come more (25%) used them for one particular parameter? I would suggest either splitting this into sub-tables, or not using all of it. For example, do you need to show the proportion using non-linear relationships for all covariate categories? The sample sizes must be small for quite a few of these.

- We agree that this table was too large and difficult to interpret. It has been replaced with Figure 4, which provides a simplified summary. Some of these categories do have low sample sizes, but this should be more apparent now as the leftmost column indicates covariate frequency.

L346: Not clear what is species specific. Do you just mean directly collected data?

- Correct, this point has been clarified on L258-259:
    - *"Directly collected data often represented finer-scale factors like prey species occurrence or details of habitat structure, which can be difficult to measure remotely."*

L348-350: Doesn't the problem here depend also on spatial scale? Satellite data may be better when sites are larger.

- This reference was removed while simplifying the results.

L364-5: Have people tried using splines?

- We did not encounter any in our sample, but now mention them as a possible subject of future research alongside machine learning methods on L508-509.

    "While no studies in our samples used these approaches, other means for flexible response shapes such as splines, BRTs, or machine learning based methods may offer other avenues for increased flexibility…"

L374-377: Frankly, I think the emphasis on more complex responses is simply because the software has been written to do it.

- This does seem plausible, and we have suggested as such on L501-502:
    - *"Many common approaches for SDMs, such as MAXENT and Boosted Regression Trees, permit considerable flexibility in the shape of their covariate response curves and use of interactions, where supported by the data (Elith et al., 2008; Merow et al., 2013)."*

Figure 5: How do these numbers relate to the amount of data? 15+ covariates is ridiculous, unless there are hundreds of sites.

- This figure includes only the number of covariates considered, not included in final models.
- We have added a scatterplot to the Supplementary material comparing the number of sites with the number of covariates considered. This indicates that a higher number of sites was associated with a higher number of covariates considered (or some summary).

L396-8: Does this matter? If the dynamics are what are being modeled, the initial occupancy isn't important, and will largely be determined by the subsequent occupancy patterns.

L398-401: If the software is written correctly, subsequent occupancy will influence the estimate of initial occupancy.

- We now restrict these concerns to cases where spatial predictions are necessary, and agree that it may not be necessary when trend estimation is more pertinent. On L450-453:
    - *"Where spatial predictions are of interest, it is prudent for users of DOMs to consider sufficient covariates to capture the key drivers of occupancy in the first primary occasion – this may be less of a concern in cases where estimating trends in occupancy probability is a greater priority."*
- The impact of unmodeled variation in occupancy, colonization, and extinction probability among sites has not been explored, a point made in both MacKenzie et al., 2006 and updated version from 2020 (L473-477).
    - *"The consequences of unmodelled heterogeneity on model outputs generally remains an understudied aspect of DOMs, despite ongoing acknowledgements that this is an area meriting further research (MacKenzie et al., 2006; MacKenzie et al., 2017). Simulation studies exploring these scenarios would help to inform where this may be of concern and guide future research on model building practices."*

L402-5: This could also raise the question of whether too many covariates are being used to model colonization and extinction.

- We note that we only recorded the number of covariates considered, and that because approaches to model selection differed, this does not always reflect how many covariates are being simultaneously included in any one model. This is explained at L273-276:
    - *"The size of the covariate pool for each parameter varied substantially, with the number of covariates considered ranging from 0 (effectively modelling the parameter as a constant) to over 40 candidates on a single parameter (Figure 5). Note that this does not represent the number of covariates included in the final model formulation used for inference."*
- We now comment on potential overfitting of models, and how appropriate model evaluation (which is currently very rarely done) can help to identify these issues (L525-258).
    - *"While overfitting models can be a concern, many covariate selection methods (including AIC-based approaches and Bayesian regularisation) will penalise excessive complexity and appropriate model evaluation (currently rare – see below) can detect these issues."*

Table 2: Did nobody use regularization? And did any of the Bayesians use variable selection methods within their MCMC? With so many potential covariates, this would seem to be a natural approach.

- No articles in our sample used regularisation, although we add a discussion of the possibility in L588-589:
    - *"range of promising methods including regularisation priors and reversible-jump MCMC merit future research on their suitability for use with DOMs (Park & Casella, 2008)."*
- One did use an approach within the MCMC, which we now mention in L314-315:
    - *"Finally, Brown et al. (2014) capitalised on the advantages of the Bayesian framework and used reversible jump MCMC routines to perform model selection during model fitting."*

Fig 6A: Would this be better as a line graph? That would make the trends easier to see.

- While we did attempt this, the resulting graph was too busy to be easily interpreted.

L534-556: I was hoping for something a bit less arm-wavy here. The general issues are well known (the authors really need to mention over-fitting somewhere, though), so the complex nature of occupancy models needs to be put into this context. My intuition is that uncertainty in detection will push towards simpler models (because the data won't support anything complex), but that will depend on a host of other factors.

- The different parts of this argument have been separated into different categories. We discuss the influence of covariates on detection in L463-472, and discuss response shapes in greater detail in L499-512.

- We now mention overfitting explicitly in L522-528:
    - *"This is not to suggest that modellers should consider arbitrary non-linear responses or covariate interactions that are not grounded in ecological theory, but that they should remain open to more complexity where merited by the data and study system. While overfitting models can be a concern, many covariate selection methods (including AIC-based approaches and Bayesian regularisation) will penalise excessive complexity and appropriate model evaluation (currently rare – see below) can detect these issues."*

L557-564: Again, can you be more concrete? Is the problem that coverage selection methods haven't been implemented in DOMs? Or that they have but aren't being used? Or that the performances of different methods in DOMs aren't well understood? Or that the reasons for using different strategies aren't clear?

- We recognise that this is one area which required much more detail and concrete recommendations. We've fully rewritten this section of the discussion, now under the "Covariate selection" subheading (L530-594). We are now more concrete about the reasons for these issues (e.g. computational constraints on exhaustive model selection, lack of studies assessing performance of different methods in DOMs), as well as the specific problems that have been identified and proposed or potential solutions.
- We first highlight why model selection is particularly challenging for DOMs and why exhaustive model selection methods may not be feasible, noting the lack of research focused on DOMs specifically. We then discuss existing research on AIC based methods and occupancy models and note some observed challenges and proposed recommendations when using these methods (L545-556):
    - *"In a comparison of static occupancy models fit with three AIC-based approaches using simulated data, Doherty et al. (2012) found that while each method achieved similar predictive performance, estimates of covariate weights varied. They advise model averaging to mitigate this effect, but acknowledge that this effect is likely to be even larger in more complex models such as DOMs."*
    - *"More recent work by Morin et al. (2020) using occupancy models fit to field data had similar findings, with sequential model selection approaches often failing to identify the lowest AIC model amongst exhaustive combinations. They recommend that modellers carry more candidate models through each stage in sequential selection processes to increase the probability that the top models are identified, given that exhaustive model selection may not be feasible for DOMs with moderate-large numbers of potential covariates."*
- We then discuss critiques of some methods for inference, which was the most comment objective in our reviewed studies, pointing readers to further discussions and highlighting tools that may be useful (L557-560):
    - *"In recent years, work from the causal modelling community has critiqued aspects of covariate selection in cases where the principal research objective is to test pre-defined environmental relationships…"*
- We end by highlighting uncertainties in Bayesian methods and highlight this as a research priority (L585-594):
    - *"Regardless of objective, Bayesian model selection for DOMs remains somewhat underexplored…"*
- We also note uncertainties in the best approaches for prediction (L580-584):
    - *"Prediction remains somewhat underexplored for DOMs, and further research is needed on the best techniques for fitting models for this purpose (Briscoe et al., 2021). While we discuss model evaluation in greater detail later in the discussion, model selection by cross-validation is a promising avenue where computationally feasible and is available via the R packages ubms and unmarked (Kellner et al., 2022, 2023; Yates et al., 2023)."*

REVIEWER 2'S COMMENTS

GENERAL

In this manuscript, the authors reviewed aspects of dynamic occupancy models that have been applied with field data and assessed covariate information, study objectives, model evaluation and performance, and

synthesized some future research directions. Overall, the manuscript had some interesting insights and the writing was fairly good. However, I feel the study could be improved with several general revisions, and more minor specific changes.

- Thank you for the thoughtful comments.

My general (high-level) recommendations:

• I suggest refining research objectives for the review to be more specific (what were the specific purposes of the review?). I agree the assessment of published dynamic occupancy models can offer some interesting insights, but these were fuzzy and many of the authors' insights were basic and could be made without such an in-depth review. I suggest specifying the review study objectives more clearly to structure what information is useful for readers.

- The objectives have been rewritten around several focal areas, with corresponding changes to the methods and results. The new objectives highlight the focus of the literature review component (L71-82):
    - *"We first identify the research contexts where DOMs have been used, characterising the research objectives they have been applied to, the scale and characteristics of the study systems where data was collected, and the methods used to collect and format the data used to fit models. We then review approaches taken to the modelling process, including the nature of the covariates considered for inclusion and the form of their relationships in the model, the methods used for model selection, and the reporting of model assessment and evaluation."*
- We then clarify that these findings are used to inform recommendations:
    - "*By jointly considering the research contexts to which DOMs are now applied and the approaches taken for model building and evaluation, we aim to highlight challenges in building DOMs, providing recommendations where possible and identifying priorities for future research where uncertainties remain."*

Likewise, the recommendations of future research are often not specific enough to provide clear direction to readers. What are the key new extensions that would improve the application of these models or solve common issues? Finally, I do not think the review would qualify as a systematic review in the sense that there was not a specific research question.

- We now focus on 5 focal areas in the discussion, and each includes at least one recommendation and at least one research priority. These are restated in a newly added summary (Table 2).
- Throughout the text we now provide more detail on research gaps, drawing on existing studies (including from occupancy modelling more generally) to outline these issues in more detail. For example (L538-542/L545-548):
    - *"Limited research has been conducted on the advantages of different methods for covariate selection in DOMs…"*
    - *"In a comparison of static occupancy models fit with three AIC-based approaches using simulated data, Doherty et al. (2012) found that while each method achieved similar predictive performance, estimates of covariate weights varied…"*
- We do not necessarily focus on extensions, as our priority is practices in implementation. One exception is variants of DOMs to support additional complexity for prediction, as mentioned on L508-512:
    - *While no studies in our samples used these approaches, other means for flexible response shapes such as splines, BRTs, or machine learning based methods may offer other avenues for increased flexibility (Hutchinson et al., 2011; Joseph, 2020; Rushing et al., 2019)."*

- We also note the need for better goodness of fit tests specific to DOMs (L599-602):
    - *"This is perhaps because model assessment methods and guidance remain underdeveloped for DOMs: few goodness-of-fit tests have been proposed, and those which do exist require more empirical assessment for greater acceptance."*
- We agree that the review is not systematic and no longer refer to it as such.

• *Simplify and shorten the text in the results. This section is long, largely narrative based and often vague ("some studies did a, some studies did b"). Restructuring based on more specific objectives should allow the Review results to be more concisely presented.*

- Thanks for this suggestion. We have returned to a more conventional format with discrete results and discussion sections. The results have been rewritten under six subheadings corresponding with the objectives and methods:
    - Research objectives
    - Study systems
    - Observations and data collection
    - Covariates and complexity
    - Model selection and assessment
    - Temporal trends in applications of DOMs

• *The writing could be improved in several places, including the specific objectives, the mathematical descriptions of the models, and when describing what exactly the results are.*

- We agree that these areas in particular required improvement.
- The objectives have been fully rewritten (L71-82).
- We found that the mathematical definitions were easier to describe in a box separate from the main text (Box 1). This allows us to do a more introductory explanation of the models, while permitting more advanced users of DOMs to skip past it without interrupting the flow of the main text.

SPECIFIC

Throughout: I found the use of "The DOM" and "DOMs" confusing and suggest revising to consistently use singular (i.e., the dynamic occupancy model family; DOM) or plural (i.e., dynamic occupancy models; DOMs) form throughout.

- We now consistently use 'DOMs' throughout.

Abstract: I don't think the models "explicitly accounts for observation error", but "imperfect detection". Observation error would also include misidentification that causes false positives and other issues. There are some extensions that can accommodate parts of these but by definition I think DOMs are primarily for dealing with imperfect detection.

- We agree that this is more precise, it has been changed.
    - *"…framework for estimating species occupancy across space and time while accounting for imperfect detection."*

L2: I think an explicit definition of "occupancy" would be helpful, given how much your review relies on a clear understanding of the term.

- We now provide a general definition of occupancy on L9-11:

    - *"Irrespective of its importance, 'occupancy' – broadly referring here to the presence of a species in a given area and time period – has proven to be a persistently difficult quantity to estimate in practice."*
- We also include a longer discussion on the nuances of defining occupancy under the subheading 'Defining occupancy and detection' (L354-361):
    - *"Building and interpreting DOMs requires a firm understanding of what probabilities of occupancy and detection conceptually represent for a given dataset. However, arriving at appropriate definitions for these terms can be a challenge. What site occupancy describes is dependent on a variety of factors, including the spatial extent of sites, the duration of primary occasions, the survey methods used, and the ways that each of these relate to the life history traits of the target species (Valente et al., 2024). Applications of DOMs in our review used data that varied across each of these axes, resulting in vastly different conceptions of occupancy and detection)."*

L56-57: I think more precise language than "at any point in time" would be helpful. Dynamic occupancy models work with discrete time so it actually estimates (predicts) occupancy at discrete capture event times, which are typically called primary surveys.

- Changed in Box 1:
    - *"Occupancy is estimated at independent sites (i) during discrete, time-bound intervals termed primary occasions (t)."*
- We have also replaced all references to 'season' with 'primary occasion' to avoid any confusion.

L63-64: You use "local" when describing extinction, but not colonisation. However, arguably both are "local".

- This was intended to reflect the original definition from Mackenzie 2003, although in the revised version we now refer to only 'colonisation' and 'extinction.'

L64: I think it would be easier to follow using ψi because you're describing in text "in following time steps". The same applied to the extinction and gamma estimates as they can be time specific in the generalized form.

- Answered below with rest of Box 1.

L67: The symbol for detection probability looks to be a special character, but an italicized "p" (p) is the typical notation for this parameter.

- Addressed below with rest of Box 1.

L68: You do not state what "y" is. Based on your description, and common notation, y = the observed detection history (i.e., a matrix of 0's and 1's for detect/non-detect at each site and for each event).

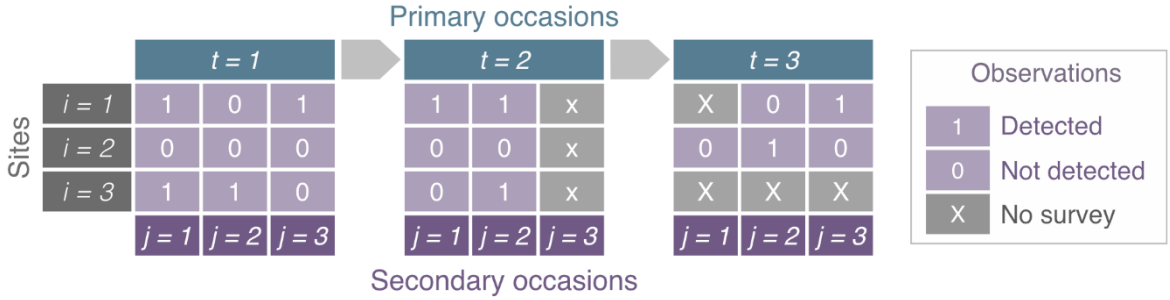
- Addressed below with rest of Box 1.

L54-74: I found the description of the model form not in an intuitive order (i.e., building complexity) and that it was missing several key definitions. For example, you described the distributions and form for deriving occupancy probability, but not colonization or extinction probability. You did not describe the links (i.e., that initial occupancy, colonization, extinction, detection are all usually on a logit link with covariates). It also does not include notations for separating sites (typically j total surveys, t primary surveys, i or k sites).

- We have moved the model introduction to a box independent of the main text to allow more flexibility in explaining the structure of DOMS.
- In the new format, we have made several changes to the notation and text to improve clarity and consistency with the broader literature:

- We have replaced all of the unique detection symbols (rho) with the more conventual *p.*
- Formulas for gamma and epsilon are still represented as constants for simplicity, but we now clarify this in Box 1:
  - *"While we have represented each of these parameters as constants for simplicity, they are more often modelled as time-dependent or estimated with respect to covariates to describe variation in the occupancy or detection processes."*

- We have changed the graphic used to represent the data used in DOMs in the first part of Box 1 to align directly with the observation data matrix y.



-
- We now define y in Box 1:
  - "For each observation, a species is either detected ($y_{i,t,j}$ = 1) or not ($y_{i,t,j}$ = 0), resulting in a three dimensional detection matrix…"

L100: Do the Miller references require "D." at the beginning?

- The D. Miller references were required due to a later citation of K. Miller; we no longer cite this paper and the 'D' has been removed accordingly.

L101: I would describe the assumption as "…closed to changes in occupancy status within seasons". By definition, occupancy can change between seasons in the dynamic occupancy model family.

- Thanks, this was a typo and has been fixed in Box 1:
  - "sites are closed to changes in true occupancy state within each primary occasion"

L136-138: I recommend adding more context to your study objectives beyond completing a systematic review on the models. What were the key questions you asked, or hypotheses you sought to test?

- We have added a dedicated objectives section in L71-82, highlighting the key areas we targeted in the literature review component and more explicitly linking them to the recommendations we make in the discussion.
- This review was somewhat exploratory and was more designed to identify key areas of interest, so we do not declare any pre-defined hypotheses.

L153-160: Based on these criteria, your review would also include papers that used similar data to model implicit-dynamics occupancy models.

- The implicit-dynamics models would not be included due to our third criteria ("had multiple seasons, between which sites could change states condition on the prior season's occupancy state and transition probabilities such as colonisation and extinction"). The criteria are more clearly stated in the revised methods section (L87-93):
  - *"Included data from multiple primary occasions, between which sites change states conditional on the prior occasions occupancy state and transition probabilities such as colonisation and extinction."*

L173: Did you use a standard taxonomic level for identifying "type of organism"?

- We did not, as we anticipated mostly vertebrate taxa and believed that *ad-hoc* categories would be more informative in the context of occupancy modelling.
- Our categories were mammals, birds, herptiles, invertebrates, and 'other'; which is now stated on L126-128:
    - *"Details collected on target taxa included the total number of taxa modelled, their general categorisation (birds, mammals, herptiles, invertebrates, or 'other'), and their conservation status."*

L186-188: Please remove the double brackets (i.e., use ";" to separate text in brackets from the reference).

- Corrected.

L189: Change "A priori" to "a priori" (italics).

- Corrected.

L195: I recommend moving the data reference to your results section below (e.g., "A total of 92 articles were included in our review (Supplementary material).")

- Implemented on L165-170.
    - *"92 articles were included for this review, fewer than the 100 possible articles due to a deficit of qualifying papers in the first stratum (2004-2007). All review data is available in the supporting information."*

Figure 4: I like the utility of this figure. However, for several (perhaps all) panels, knowing the total sample size (92) is helpful. Consider trying to incorporate into part of the figure to make the denominator/total explicit.

- This was an important note, as the sample sizes are not all equivalent. We have added the appropriate sample size to each figure, and provided additional details in the caption of Figure 3:
    - *"A summary of the research contexts in reviewed applications of DOMs. a) The approximate locations of the study areas where data was collected. b) Spatial extent of the study areas, defined as the area of inference within which all surveyed points were contained. c) Number of articles which fit models to each taxa category. Taxa are considered 'threatened' if they are listed on the IUCN Red List, or if authors indicate that they are otherwise threatened. d) The number of taxa modelled as distinct groups in each application. Explicitly multi-species models include both hierarchical jointly estimated models and more interactive models. e) Number of articles using each method for capturing detection/non-detection data. These are non-exclusive, and citizen science data is a subset of 'direct observations.' (f) The number of sites and study duration for each dataset used to fit DOMs. Gold lines indicate medians."*

L207: What is "all geographic realms"? All continents, except Antarctica? You state "broad diversity of ecosystems", but the breadth, list, or diversity of ecosystems is not included in Figure 4 (A or otherwise).

- We have removed these mentions on (L196-197), instead referring to continents:
    - *"study systems spanned all continents except Antarctica"*

L246-254: It would be useful to have numbers (e.g., % of articles, or # of articles) to put "some" in context here.

- This line has been removed when simplifying the results.

L257-259: Earlier you reported that multi-species models were common with acoustic data, but don't include an acoustic sensor category (e.g., Autonomous Recording Units) similar to camera traps. Were none included in your review?

- By chance none were included in our sample – we instead discuss their suitability in L58-61:
    - *"Technological advances have facilitated the widespread deployment of autonomous detection methods including camera traps and acoustic monitors, generating large quantities of observation data suitable for analysis with DOMs (Balantic & Donovan, 2019; Lahoz-Monfort & Magrath, 2021)."*

L279-280: Here and elsewhere you slip into describing articles or studies in the present tense, but all the work is in the past and should be described in the past tense.

- Tense has been corrected throughout.

L308: Change to "…at an extremely fine temporal scale…"

- This line has been removed.

L309-311: Can you be more specific about what research questions could be used to provide general guidance on season length?

- This is now discussed in L416-427 of the 'Defining occupancy and detection' section:
    - *"An unusual challenge arises when users seek to fit to DOMs to detection/non-detection data that is continuously recorded – this may occur with camera traps or acoustic monitors which can run for long time periods, or with data from citizen science portals that lacks clearly delineated survey periods. In these situations, modellers must arbitrarily determine how to stratify detections into primary and secondary occasions (but note the availability of continuous detection models for secondary "occasions"; (Emmet et al., 2021; Guillera-Arroita et al., 2011; Pautrel et al., 2024). This stratification affects the temporal scale at which occupancy is measured, with corresponding changes to the definitions of occupancy and detection. The appropriate duration of these occasions will of course depend on research objectives and the ecology of the target species (Chave, 2013), but guidelines for approaching this task in DOMs specifically would hold value for the increasing numbers of authors fitting models to this type of data."*

L327: Can you elaborate on what "site geometry" is?

- Fixed on L245-246:
    - *"35% of studies incorporated covariates for site geometry and connectivity, such as habitat patch size or distance to other sites"*

L335: Use consistent capitalization (e.g., "Spotted owl" vs "Barred Owl") and usually you should state the scientific name the first time a species is named.

- Fixed on L177-178.
    - "influence of barred owls (*Strix varia)* on the threatened spotted owl (*Strix occidentalis*)."

L350: Remove the word "percent".

- Fixed.

L386-388: This conclusion assumes that all considered covariates were reported in the manuscript. Many studies likely explored additional covariates but don't report on them and it may be worth noting this with a nod to publication bias (i.e., the final versions are described, perhaps with some model selection, but that it doesn't include every covariate coauthors considered).

- This is a good point that we now acknowledge on L144-146:
    - *"we recorded all covariates considered in each study regardless of whether they were included in final models, acknowledging that not all covariates considered are always reported"*

L472: Replace "that can been" with "that has been".

- This line has been removed.

L513,518: See my earlier comments about the use of English common names and capitalization.

- Fixed on L180:
    - *"… habitat covariates for wood warbler (Phylloscopus sibilatrix) occupancy"*

Figure 6: The caption here (and in other figures) should be more independent for standing alone. For example, include that the context of objectives is for dynamic occupancy models.

- The captions for all six figures have been rewritten to stand independently. E.g. Figure 6 caption now reads:
- *Figure 6: Key trends in applications of dynamic occupancy models over the study period. a) The number of sites modelled in each dataset, presented on the log scale. b) The number of covariates considered per model parameter in each modelling workflow, presented on the log scale. c) The percentage of models in each stratum fit in the maximum likelihood or Bayesian frameworks. d) The percentage of articles in each stratum that used each of four non-exclusive covariate selection strategies. This panel includes only articles which fit models via maximum likelihood. e) The percentage of articles in each stratum that reported goodness-of-fit testing and model validation (with either in or out of sample data).*

L564-569: Part of the lower number of studies in your sample that didn't evaluate model performance or goodness-of-fit is likely because of when dynamic occupancy model goodness-of-fit tests were deployed in common R packages compared to the sample of time in your study.

- True, changes in this rate are now indicated by the new time-series Figure 6. Even in the most recent strata, model performance and goodness-of-fit testing remain rare (although they have improved) (L338-341):
    - *"While there is evidence of improvement from earlier strata, goodness-of-fit testing and model validation using either in- or out-of-sample data remained uncommon even in more recent strata where tests were readily available in common R packages."*