Dynamic occupancy models for applied ecology: a review and framework for implementation

Target Journal: Ecography

# Abstract

Dynamic occupancy models are a key tool for ecologists seeking to understand patterns of occupancy across space and time while accounting for imperfect detection. Thanks to their relative ease of implementation, realistic data requirements, and capacity to generate useful estimates of occupancy parameters, these models have become popular in modelling natural systems for conservation and wildlife management objectives. We conducted a systematic review of studies incorporating these models to assess what they are used for and how they are implemented in applied studies. Our findings indicate these models experience substantial use in assessing occupancy trends and in evaluating specific relationships for a great diversity of taxa, but are less frequently used to generate either spatial or future predictions. However, their implementation has not always been consistent – decisions in the model-building process such as covariate inclusion, model selection, and model evaluation are highly variable between authors, with potential implications for upholding model assumptions and the robustness of outputs. To this end, we provide some guidelines for future authors seeking to implement dynamic occupancy models to ensure that key considerations are accounted for in the model building process.

**Ecography author guidelines:**

**Review and synthesis papers** provide a critical assessment of the literature with emphasis on current topics in which rapid and significant advances are occurring. Items in this category should be more focused than the broad, topical reviews typically published elsewhere, developing a synthesis that inspires new hypotheses or new methods. Contributions in this category will be solicited by the editors. However, unsolicited submissions will also be considered and sent for pre-submission assessment by Review & Synthesis editors. Review articles should be maximum 7500 words in length (main text) and maximum 10 figures, tables or boxes.

# Introduction

## Overview

Capturing patterns of species occupancy over space and time is a common goal for ecologists, particularly those focused on conservation and wildlife management. Advances in recent decades have provided numerous options for methods and statistical models, with sub-fields such as species distribution modelling (Elith & Leathwick, 2009) and metapopulation modelling contributing increasingly sophisticated tools to support on-ground practitioners. Of course, no matter the method ecologists must balance data input requirements and analyst skillsets against inferential power and fit to purpose when determining how to best analyse data from natural systems.

MacKenzie et al.’s 2002 paper, ‘Estimating site occupancy rates when detection probabilities are less than one,’ (MacKenzie et al., 2003) first defined what is now termed the ‘Dynamic occupancy model[[1]](#footnote-1)’ (henceforth DOMs - see ‘Box 1’ for details on the basic model structure). The model is a well-balanced option for many applied ecologists: it requires achievable presence/absence counts; albeit with revisits during each primary sampling occasion; yet provides estimates of initial occupancy, colonisation, extinction, and detection probabilities which capture aspects of processes not included in other alternatives (Bailey et al., 2014).

Box 1: What are dynamic occupancy models?

**Model definition**

The basic structure of the model is simple, consisting of an occupancy module and an observation module. In the occupancy module, independent sites may exist in either occupied or unoccupied states; transitions between the two between time steps are termed colonisation and extinction. In the observation module, we account for imperfect detection by conducting multiple surveys within a single timestep.

**[GRAPHICAL REPRESENTATION HERE]**

**[Key formulas to go here]**

**Assumptions**

1. Sites are considered ‘closed’ between time-steps, with occupancy state presumed to be un-changed.
2. There are no false positive detections.
3. No unmodelled heterogeneity exists.

**Model extensions**

Beyond the basic form of the model, authors have contributed several other extensions to DOMs. Most popular are the multi-species extensions, valued for modelling interspecific relationships, and multi-state options with use for incorporating demography among other variations in what constitutes occupancy. Also available are variants accounting for false positive error.

Researchers in governmental agencies, academic institutions, and non-governmental organisations (NGOs) have implemented DOMs for a wide range of species and purposes, from estimating occupancy patterns of threatened species to monitoring the range expansion of invasive species (Broms et al., 2016). Since its publication, MacKenzie et al.’s 2002 model-defining paper has been cited 4962[[2]](#footnote-2) times increasing year-over year – a testament to their continued popularity within the applied modelling community.

DOMs do have their pitfalls, however. Their assumptions (see Box 1), while relatively straightforward, are also unlikely to be entirely fulfilled in natural systems. While a reasonable amount of work has been conducted on the importance of the closure assumption (Otto et al., 2013; Rota et al., 2009), comparatively little has been done with respect to the heterogeneity assumption (MacKenzie et al., 2017). Heterogenous landscapes and patterns of occupancy are the norm in ecological systems and fully satisfying this assumption is not realistic in natural systems. The main mechanism for users to address this concern is by including covariates in models to estimate parameters to describe how environmental variation may affect probabilities of occupancy or detection.

Anecdotally, the quality and quantity of covariates can range widely depending on system, data availability, and accessible computational resources. The matter in which models are implemented, including the data inputs, covariates considered, manner of final model selection, and evaluation of model fit are all important contributors to the reliability of model outputs. When conservation and management decisions are made based on model outputs, it is critical to ensure that all steps are fully considered to develop the best possible model.

This paper has two principal objectives:

1. Assess how DOMs have been used in applied ecological research, including the types of projects in which practitioners used them and how they have implemented them, with emphasis on the model selection process.
2. To provide practical recommendations for how to use DOMs to ensure outputs and predictions are as robust as possible, with a practical workflow for development to incorporate key considerations.

# Review methods

## Paper elicitation

This review was focused only on papers which model occupancy across space and time using real data. To quality for inclusion, papers were required to fulfil each of these criteria:

* Multiple sites capable of exhibiting two or more occupancy states; including at least an occupied and unoccupied state.
* Multiple time-steps between which occupancy states can change, with transitions between states modelled as Markovian processes.
* Data collected from a natural system, not theorical or simulated. The data need not have been explicitly collected for the given paper.

Four search terms[[3]](#footnote-3) were used to generate the initial pool of papers:

* Dynamic occupancy model
* Occupancy dynamics model
* Multi-season occupancy model
* Stochastic patch occupancy model

Each term was searched on Google Scholar (Appendix I), which uses a proprietary ranking system based on relevance. The first 100 results (if available) for each term were considered for inclusions. Non-English papers, those clearly outside the field of ecology, or those not accessible via Google Scholar or the University of Melbourne library were immediately discarded. 287 papers remained for consideration at this stage.

## Preliminary and formal reviews

The pool of papers was stratified by search term and publication period[[4]](#footnote-4) and randomly ranked within their strata. Papers in the lowest 25% or lowest 5, whichever was larger, were marked for inclusion in review. In cases where papers did not meet qualification criteria, they were replaced by the next lowest paper in their strata where available.

Authors developed a structured spreadsheet with categories for study metadata, objectives, taxa, location, survey methods, detection, covariates, modelling, and outputs. Findings were systematically noted as each paper was read; 75 papers were included at this stage.

Study questions were further refined after the preliminary review, and a revised spreadsheet with more clearly articulated parameters was generated (Appendix II). The authors also determined that ‘Stochastic patch occupancy models’ represented a distinct model form from the other three search queries, with a unique history and distinct qualities. Therefore, we decided to exclude these papers (n = 21) from the formal review.

For the formal review, all remaining papers were re-read and their results logged in the spreadsheet. The final review spreadsheet is presented in Appendix II. The final count of qualified papers was n = 54. All subsequent analyses were conducted in R.

# Results

**Authorship and location**

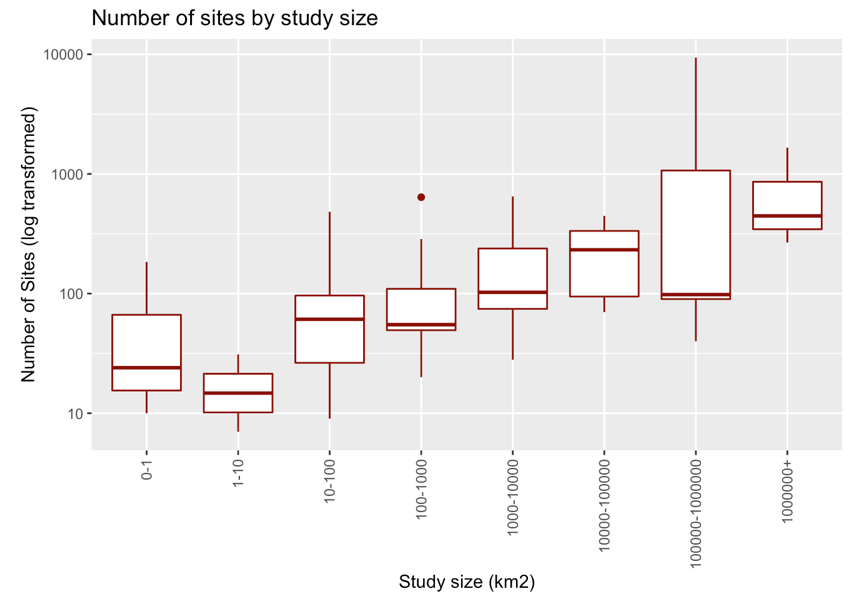
Following general trends in ecological publishing (Nuñez et al., 2021), an outsized share of publications included in the review were published by first authors based at United States institutions, with no other country producing more than 4 papers to the U.S.’s 36 (Table I). Similar trends are present in the locations of study areas; five countries are represented only in the former category (note that only the first author’s institution’s location is documented). Most study sites (30/54) were situated in either the Nearctic or Palearctic realms, although papers from the Neotropical, Afrotropical, Indo-Malayan, and Australasian realms were also included.

*Table 1: Locations of authors and study areas.* ***Bold*** *indicates countries only included in the former category.*

|  |  |  |  |
| --- | --- | --- | --- |
| First author country | Number of papers (each) | Study location | Number of papers (each) |
| United States | 36 | United States | 31 |
| Australia | 3 | India | 4 |
| Argentina, Brazil, India, United Kingdom | 2 | Australia | 3 |
| Canada, France, Italy, Norway, South Africa, Spain, Switzerland | 1 | Argentina, Brazil, **Costa Rica**, South Africa, Switzerland, United Kingdom | 2 |
|  |  | **Belize**, **Cambodia**, Canada, **Eswatini**, Italy, **Lesotho**, Norway, Spain | 1 |

Author institutions were divided into four categories – academic institutions, government agencies, non-governmental organisations, and private companies. 81.5% of papers included at least one academic-affiliated author, 42.6% included a governmental agency, 29.6% an NGO, and 7.4% private for-profit companies.

**Taxa and survey methods**

DOMs in the review sample were fit to a diversity of taxa spanning mammals, birds, herptiles, invertebrates, and freshwater fish. 16/54 involved threatened species, and another 8 included invasives. As taxa generally inform the choice of survey method – the data used to generate detection histories included standardised surveys such as point counts or transects (22/54), exhaustive site searches (13/54), physical trapping (12/54), camera trapping (8/54), and one case employing bioacoustics monitors. An additional 7/54 used citizen-science data in some form.

Surveys varied substantially in their physical and temporal scale. Projects ranged from study areas from under a km2 to over a million km2 in the extent of their area of inference, with the median project falling in the 1000-10000km2 bracket. These projects contained 7 to 9394 individual sites, with an average of 377. Number of sites increased with project size (Figure 2) The timespan from the first to last data point ranged from under a month to 40 years in duration, with an average project duration of 7 years.

*Figure 1: Number of sites increases with size of study area*

**Research objectives and outputs**

Five categories of objectives and goals were assessed for each paper; these were not mutually exclusive, and many papers had multiple objectives.

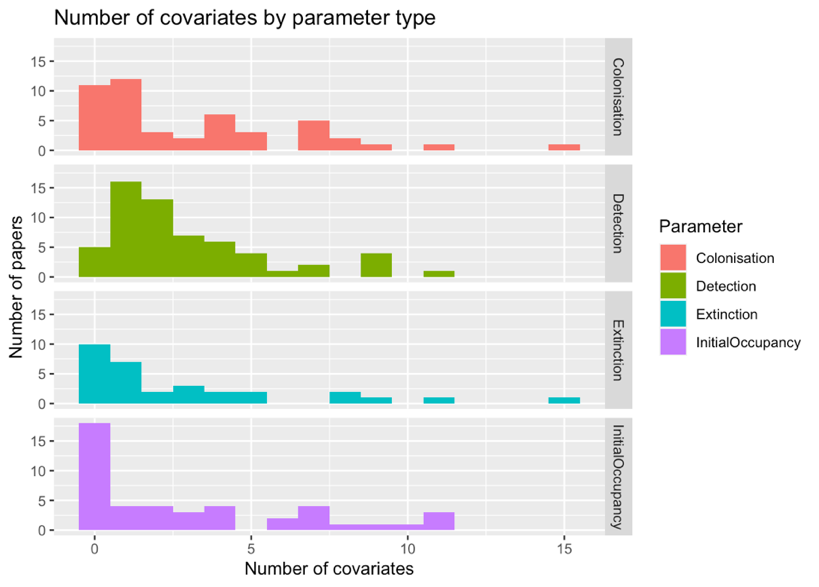
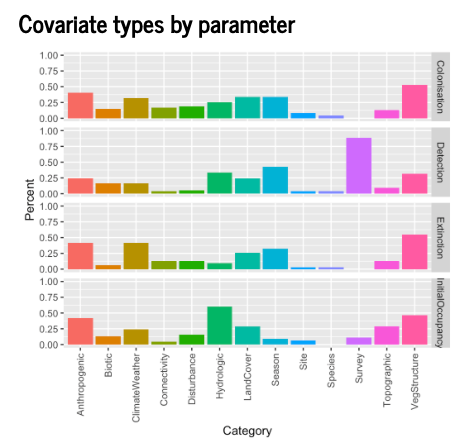
* **72.2%** were assessing trends: interested in the trajectory of estimated or derived parameters.
* **55.6%** were testing hypotheses: evaluating attributes of a **specific** covariate on parameters.
* **18.5%** were methods oriented: introducing, testing, or demonstrating a new extension to the DOM.
* **9.3%** conducted spatial mapping: Extrapolating estimates of parameter values beyond surveyed sites.
* **7.4%** made future projections: Making predictions of parameters into the future.

**Parameters and covariate inclusion**

Nearly all papers used the four core parameters of the standard DOM: Initial occupancy, colonisation, extinction, and detection (although 12 used the reciprocal of Extinction, Persistence). 9 papers included at least one multi-species parameter (one conditional on the occupancy/detection state of another species) and 8 included parameters beyond the four core parameters, such as false-positive rates or transitions to states beyond occupied and unoccupied.

*Figure 2: Covariate types (A) and quantity (B) varied with the parameter of interest.*

Covariates considered for inclusion were diverse, and varied with both the taxa and parameter of interest. Initial occupancy and detection probability generally considered fewer covariates than either colonisation or extinction did.



Surprisingly, higher-order covariations and interaction terms were rare. Just 16.7% of papers considered a quadratic term and only 18.5% of papers considered at least one interaction term.

**Model selection and evaluation**

65% of papers in the review set did not undergo a formal model selection process, using an a priori model. Of the remaining papers, the most common methods of model selections used multi-step processes such as stepwise approaches (15%) or parameter-by-parameter methods where each parameter was fixed in turn. A remaining 7% of papers used methods filtering based on simpler models, i.e., taking the covariates from the best single-season model.

Frequentist and Bayesian implementations diverged in their approach to model selection, with Bayesian implementations far more likely to define a single model a priori (Figure 3).

Chart

Description automatically generated with medium confidenceJust one Bayesian implementation used a form of model selection: Rowe et al., 2019 reduced their initial covariate set by only including covariates which surpassed a threshold of the posterior distribution mass in a preliminary model.

Overall, model evaluation (beyond AIC scores) was quite uncommon – just 5/54 papers used any such form, e.g., out-of-sample validation.

# Discussion

Since their inception dynamic occupancy models have been widely used tool to study a diverse cohort of organisms and ecosystems around the world, albeit with a strong publication bias towards western institutions. Projects which have implemented DOMs for their analysis range wildly in their scale and scope, from very small short-term projects (Button et al., 2019) to continental-scale long term monitoring programs (Weir et al., 2009; Zuckerberg et al., 2011). This flexibility and breadth of publications is an asset for DOMs, and an indication that existing trends of their increasing popularity may continue.

**Covariates**

The diversity of covariates considered for DOMs is impressive, and the heightened availability of high-quality continent scale data should provide authors with yet more options to model their systems as accurately as possible. Notably, the number of covariates considered for inclusion varies with the authors’ objectives (Figure 4), with papers assessing trends in occupancy generally using more covariates than those testing specific hypotheses. This is likely because authors in the latter category tend to focus on a specific covariate of interest, for example the occurrence of an interacting species (Mangan et al., 2019) or response to natural disaster (Falcy & Danielson, 2014). However, focusing only on one covariate at the cost of others can run afoul of the heterogeneity assumption, which purports that the key environmental drivers of variation are accounted for. In cases where this is not reasonably fulfilled variation may be misattributed to those covariates which were included, resulting in potentially misleading results.

The infrequency of higher order covariates and interaction terms is another potential concern in the covariate space. While inclusion of these terms can substantially increase the number of models which must be fit, the complexity they allow the shape of relationships to take is necessary to accurately reflect environmental realities, such as ubiquitous quadratic responses to temperature or temperature and rainfall interaction effects. As computing resources have expanded substantially since earlier DOM papers, the extra compute cost required to more exhaustively consider complex covariates may pay dividends in model fit. This is, however, an area of limited prior research and further consideration is required to determine how the omission of these terms may affect model performance.

**Model selection and evaluation**

This review found little uniformity in the methods used to select covariates for final models. Methods did vary considerably with project objective; hypothesis-testing papers were more likely to assign a model a priori than those papers which assessed trends or made predictions. This is perhaps related to aforementioned priorities around covariate inclusions and bears similar concerns with respect to the heterogeneity assumption.

While some general approaches were relatively common, such as the combination of stepwise selection with fixing the parameters one by one, details of these methods were idiosyncratic, such as the preference for null or global base models or the order in which parameters are fixed. There has been limited work conducted on how these methods differ in the final models they produce, and whether these decisions may affect model performance is unknown and an avenue for future research.

Notably, there is a large discrepancy between frequentist and Bayesian models, the latter of which were nearly always fit a priori. The cause of this disparity is unclear, but may be related to increased computing requirements. In any case, the consequences are again understudied and require further investigation.

# Conclusions

This review finds that DOMs have been heavily employed for important research with direct conservation and management implications. With increased proliferation of survey methods like camera traps and acoustic monitors which generate data suitable for DOMs, it is probable that this popularity will continue into the future. However, significant questions remain around how applied ecologists can make their models as robust and accurate as possible. Key decisions around the modelling, including covariate inclusion and model selection, are applied highly idiosyncratically, with little research available to inform these choices. These areas of concern should be a priority for future research to ensure that decisions made based on results of dynamic occupancy models are as reliable as possible.

# References

Astete, S., Marinho-Filho, J., Kajin, M., Penido, G., Zimbres, B., Sollmann, R., Jácomo, A. T. A., Tôrres, N. M., & Silveira, L. (2017). Forced neighbours: Coexistence between jaguars and pumas in a harsh environment. Journal of Arid Environments, 146, 27–34. https://doi.org/10.1016/j.jaridenv.2017.07.005

Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models. Methods in Ecology and Evolution, 5(12), 1269–1279. https://doi.org/10.1111/2041-210X.12100

Broms, K. M., Hooten, M. B., Johnson, D. S., Altwegg, R., & Conquest, L. L. (2016). Dynamic occupancy models for explicit colonization processes. Ecology, 97(1), 194–204. https://doi.org/10.1890/15-0416.1

Button, S. T., Sovie, A. R., Greenberg, C. H., & Austin, J. D. (2019). Evaluating the Ecology of Tantilla relicta in Florida Pine–Wiregrass Sandhills Using Multi-Season Occupancy Models. Journal of Herpetology, 53(3), 179. https://doi.org/10.1670/18-112

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics, 40, 677–697.

Falcy, M. R., & Danielson, B. J. (2014). Post-hurricane recovery and long-term viability of the Alabama beach mouse. Biological Conservation, 178, 28–36. https://doi.org/10.1016/j.biocon.2014.07.010

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., & Franklin, A. B. (2003). ESTIMATING SITE OCCUPANCY, COLONIZATION, AND LOCAL EXTINCTION WHEN A SPECIES IS DETECTED IMPERFECTLY. Ecology, 84(8), 2200–2207. https://doi.org/10.1890/02-3090

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2017). Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier Science & Technology. http://ebookcentral.proquest.com/lib/unimelb/detail.action?docID=5124550

Mangan, A. O., Chestnut, T., Vogeler, J. C., Breckheimer, I. K., King, W. M., Bagnall, K. E., & Dugger, K. M. (2019). Barred Owls reduce occupancy and breeding propensity of Northern Spotted Owl in a Washington old-growth forest. The Condor, 121(3), duz031. https://doi.org/10.1093/condor/duz031

Nuñez, M. A., Chiuffo, M. C., Pauchard, A., & Zenni, R. D. (2021). Making ecology really global. Trends in Ecology & Evolution, 36(9), 766–769. https://doi.org/10.1016/j.tree.2021.06.004

Otto, C. R. V., Bailey, L. L., & Roloff, G. J. (2013). Improving species occupancy estimation when sampling violates the closure assumption. Ecography, 36(12), 1299–1309. https://doi.org/10.1111/j.1600-0587.2013.00137.x

Rota, C. T., Fletcher Jr, R. J., Dorazio, R. M., & Betts, M. G. (2009). Occupancy estimation and the closure assumption. Journal of Applied Ecology, 46(6), 1173–1181. https://doi.org/10.1111/j.1365-2664.2009.01734.x

Rowe, J. C., Duarte, A., Pearl, C. A., McCreary, B., Galvan, S. K., Peterson, J. T., & Adams, M. J. (2019). Disentangling effects of invasive species and habitat while accounting for observer error in a long‐term amphibian study. Ecosphere, 10(4). https://doi.org/10.1002/ecs2.2674

Vasudev, D., Goswami, V. R., & Oli, M. K. (2021). Detecting dispersal: A spatial dynamic occupancy model to reliably quantify connectivity across heterogeneous conservation landscapes. Biological Conservation, 253, 108874. https://doi.org/10.1016/j.biocon.2020.108874

Weir, L., Fiske, I. J., & Royle, J. A. (2009). TRENDS IN ANURAN OCCUPANCY FROM NORTHEASTERN STATES OF THE NORTH AMERICAN AMPHIBIAN MONITORING PROGRAM. Herpetological Conservation and Biology, 15.

Zuckerberg, B., Bonter, D. N., Hochachka, W. M., Koenig, W. D., DeGaetano, A. T., & Dickinson, J. L. (2011). Climatic constraints on wintering bird distributions are modified by urbanization and weather: Wintering birds, weather, food, and climate. Journal of Animal Ecology, 80(2), 403–413. https://doi.org/10.1111/j.1365-2656.2010.01780.x

1. Also variously termed ‘occupancy dynamics models’ and ‘multi-season occupancy models [↑](#footnote-ref-1)
2. Google scholar citation figure [↑](#footnote-ref-2)
3. Plus grammatic variations [↑](#footnote-ref-3)
4. 2000-2005, 2006-2010, 2010-2015, 2015-2021 [↑](#footnote-ref-4)