A review of dynamic occupancy models and framework to building them for applied ecology

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 significant use in assessing occupancy trends and in evaluating specific relationships for a great diversity of taxa, with far more limited use in generating 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 robustness of outputs. To conclude, we provide some simple guidelines for future authors seeking to implement dynamic occupancy models to ensure that key considerations are accounted for in the model building process.

# 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 and metapopulation modelling contributing increasingly sophisticated tools to support on-ground practitioners. No matter the method, ecologists must balance data input requirements and analyst skillsets against inferential power and suitability to purpose when determining how best to analyse data from natural systems.

MacKenzie et al.’s 2002 paper, ‘Estimating site occupancy rates when detection probabilities are less than one,’ first defined what is now termed the ‘Dynamic occupancy’ model[[1]](#footnote-1) (henceforth DOMs). 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. These modules combine to provide valuable estimates of initial occupancy, colonisation, extinction, and detection probabilities.

DOMs operate under just three key 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.

Of course, heterogenous landscapes and patterns of occupancy are the norm in ecological systems – to satisfy this third assumption, covariates may be used for each model parameter to describe how environmental variation may affect probabilities of occupancy or detection.

MacKenzie et al.’s 2002 model defining paper has been cited 4962[[2]](#footnote-2) times, with publications increasing year-over year since their inception. It is likely that this is because DOM’s strike a balance between ease of use, realistic data requirements, and assumptions with analytical capabilities suitable for answering many types of questions common in ecology and conservation science.

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 on the basis of these models, it is critical to ensure that all steps are fully thought through to develop the best possible model.

This paper has two principal objectives:

1. To review the history of DOMs development and their use in applied ecological research, including for what practitioners used these models and how they implemented them.
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.

## Model development and history

The standard form of the dynamic occupancy models was initially described by MacKenzie et al. 2002 in their Ecology paper ‘Estimating site occupancy, colonisation, and local extinction when a species is detected imperfectly’. This title highlights one of the key benefits of the model - the ability to account for imperfect detection is a ubiquitous problem in surveying wildlife, and the subject of a great deal of discussion in this field.

Bayesian version

Multi-species extension

Multi-state extension

False positive extension

# 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 an occupied and unoccupied state.
* Multiple time-steps between which occupancy states can change, with transitions between states modelled as a Markovian process.
* Data must be collected from a natural system, not theorical or simulated. The data need not have been explicitly collected for the given paper.

Following internal discussions, 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). The first 100 results (if available) for each term were considered for inclusions, although 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 if 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 better 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 count of qualified papers was n = 54. All analyses were conducted in R.

# Results

Split out by categories:

Authorship

Taxa/Location

Objectives

Survey details

Covariates

Model selection

# Discussion

Objective x Covariate use discrimination

Concern over disparate model selection methods

Disparities in Bayesian vs non-Bayesian

Particular concern over the lack of higher order & interaction terms

Rarity of model evaluation

Guidelines in box?

# Conclusions

Highlight need for more research on what happens when the heterogeneity assumption is not fulfilled

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 variation [↑](#footnote-ref-3)
4. 2000-2005, 2006-2010, 2010-2015, 2015-2021 [↑](#footnote-ref-4)