Twenty years of dynamic occupancy models: a review of applications and look towards the future

Authors: Saoirse Kelleher[1], Gurutzeta Guillera-Arroita[2,1], Jane Elith[1], Natalie Briscoe[1]

Affiliations: [1]University of Melbourne School of Agriculture, Food, and Ecosystem Sciences; [2]Pyrenean Institute of Ecology, Spanish National Research Council (CSIC)

Corresponding author: Saoirse Kelleher (saoirse.kelleher@unimelb.edu.au)

Abstract: Understanding patterns of species occupancy across landscapes and throughout time is a long-standing objective of ecological research that has inspired the development of numerous quantitative modelling approaches. However, estimating occupancy can be a challenge, particularly when contending with issues like imperfect detection and shifting distributions. Dynamic occupancy models (DOMs) offer a framework for occupancy estimation that explicitly accounts for observation error while capturing the mechanisms driving occupancy dynamics by estimating colonisation and local extinction processes. In light of increasing interest in more process-explicit models for understanding species occurrence, here we examine how DOMs have been applied to field ecological data in the two decades since their introduction. Following a general introduction to the model, we present the results of a systematic review exploring where and how DOMs have been applied. We interrogate how authors have built their models, with particular emphasis on how covariates are incorporated to describe variation in occupancy dynamics. Our findings indicate that DOMs are a flexible tool readily applied to diverse study systems and data types, with their usage expanding in recent years as more studies apply them to make spatial and temporal predictions of species occupancy. DOMs are also amenable to extension, further broadening their utility. However, model complexity in DOMs tends to be low; most studies consider relatively few covariates and these are typically represented as simple linear relationships. Approaches to covariate selection also vary considerably, and there remains little research

on how these choices may influence model performance. Furthermore, only a fraction of articles report evaluating DOMs and little guidance exists on how to approach this task. These uncertainties in the modelling process should be key priorities for future research on DOMs given their increasing use in applied ecological research.

Introduction

The description of patterns of species occupancy across landscapes has been a long-standing subject of ecological research (Humboldt, 1849). Estimates of how widespread a species is and where it occurs are the foundation of monitoring programs and important for assessing conservation status while identifying potential drivers of occurrence can help inform potential management actions (MacKenzie & Reardon, 2013). Robust knowledge of the occupancy patterns of a species can also help us to predict where a species is most likely to occur, both under present conditions and in hypothetical future scenarios (Kéry et al., 2013).

While occupancy is a useful concept, it is also a challenging quantity to describe, measure, and estimate. The need to understand and quantitatively describe species occupancy has led to the development of several popular modelling approaches, including stochastic patch occupancy models commonly applied to study meta-population dynamics (Gutiérrez-Arellano et al., 2024), and species distribution models (SDMs) widely used to explore species occurrence at larger scales (Franklin, 2010). However, several factors can make occupancy difficult to estimate. For instance, simple presence/absence observations can be biased when species are detected imperfectly – this is often the case in wildlife field data, where it can be impossible to determine from a single observation whether a location is truly occupied or whether the species occurs but was not detected (Gu & Swihart, 2004; Lahoz-Monfort et al., 2014). Despite the ubiquity of imperfect detection in data collection, many models fit to presence/absence data make no adjustment for this source of bias (Kellner & Swihart, 2014). Another challenge for modelling occupancy is the difficulty in describing populations under non-equilibrium conditions, where a species' occurrence pattern and relationship to its environment is in flux (Dormann, 2007; Elith et al., 2010). These conditions often occur during biological invasions and climate change driven range shifts, each of which are conservation priorities and increasingly common scenarios in the Anthropocene (Bertelsmeier et al., 2013; Lenoir & Svenning, 2015).

The site occupancy models first introduced by MacKenzie et al. (2002) set the foundation for a powerful framework for modelling presence/absence data while accounting for each of these challenges (Guillera-Arroita, 2017). Drawing on principles from the mark-recapture literature, occupancy models use multiple resurveys of sites to estimate detection probabilities and correct for bias in estimates of site occupancy. Originally a static model, MacKenzie et al. (2003) extended this model for use in multiple time-steps by explicitly describing the process of changing occupancy via colonisation and extinction, relaxing assumptions of equilibrium and allowing description of patterns of site occupancy through time. This dynamic occupancy model (DOM) balances complexity and feasibility, explicitly describing the key processes driving occupancy dynamics while requiring reasonably simple-to-collect presence absence/data instead of the detailed demographic or abundance data required by more process-explicit models (Briscoe et al., 2019). These features make the DOM an important tool, with uses including hypothesis testing of relationships between occupancy and the environment, explorations of the key drivers of occupancy, and even prediction of occupancy under future conditions (Briscoe et al., 2021; Kéry et al., 2013).

In this review we examine how dynamic occupancy models have been used by ecologists in the two decades since their inception. Following an introduction to the model's form and assumptions, we present the results of a systematic review exploring how researchers have applied DOMs to ecological data, with emphasis on how they collected their data, selected covariates, and evaluated their models. Based on these results we highlight the DOM's flexibility as a tool for understanding species occurrence, examine approaches to the model building process, and outline key priorities for future research involving this model class.

Dynamic occupancy model form and assumptions

The DOM structure encompasses two processes: the ecological process of site occupancy dynamics describing the presence or absence of a species at a site at any point in time, and the observational process of detection describing whether a species is observed at a site where it is present (Figure 1). In the latent ecological process of site occupancy, sites may exist in either an occupied ($z = 1$) or unoccupied state ($z = 0$). In the first time-step, occupancy is determined by the probability of initial occupancy ψ_1, such that $z_1 \sim \mathrm{Bernouilli}(\psi_1)$. In following time steps, occupancy state can change according to a Markovian process where occupancy is predicated on the site's state in the prior time-step and the probabilities of colonisation γ and local extinction ε, such that $z_t \sim \mathrm{Bernoulli}(\psi)$ where $\psi = z_{t-1}(1-\epsilon) + (1-z_{t-1})\gamma$. Where the occupancy process is latent, the observation component of the DOM describes the probability of recording a detection when a site is occupied, allowing the DOM to account for imperfect detection. At occupied sites, the detection probability ρ describes whether the species is observed (y = 1) during a survey j or whether it is not (y = 0), according to the formula $y_{t,j} \sim \mathrm{Bernouilli}(z_t \rho)$. Under the DOM's original parameterisation, it is assumed that false-positive detections at unoccupied sites never occur. In this review we use the term "parameter" to refer to the quantities being estimated: in most cases, these are the probabilities initial occupancy ψ_1, colonisation γ, extinction ε, and detection ρ (see Figure 1). However, some alternative model forms may directly estimate occupancy ψ as a parameter.

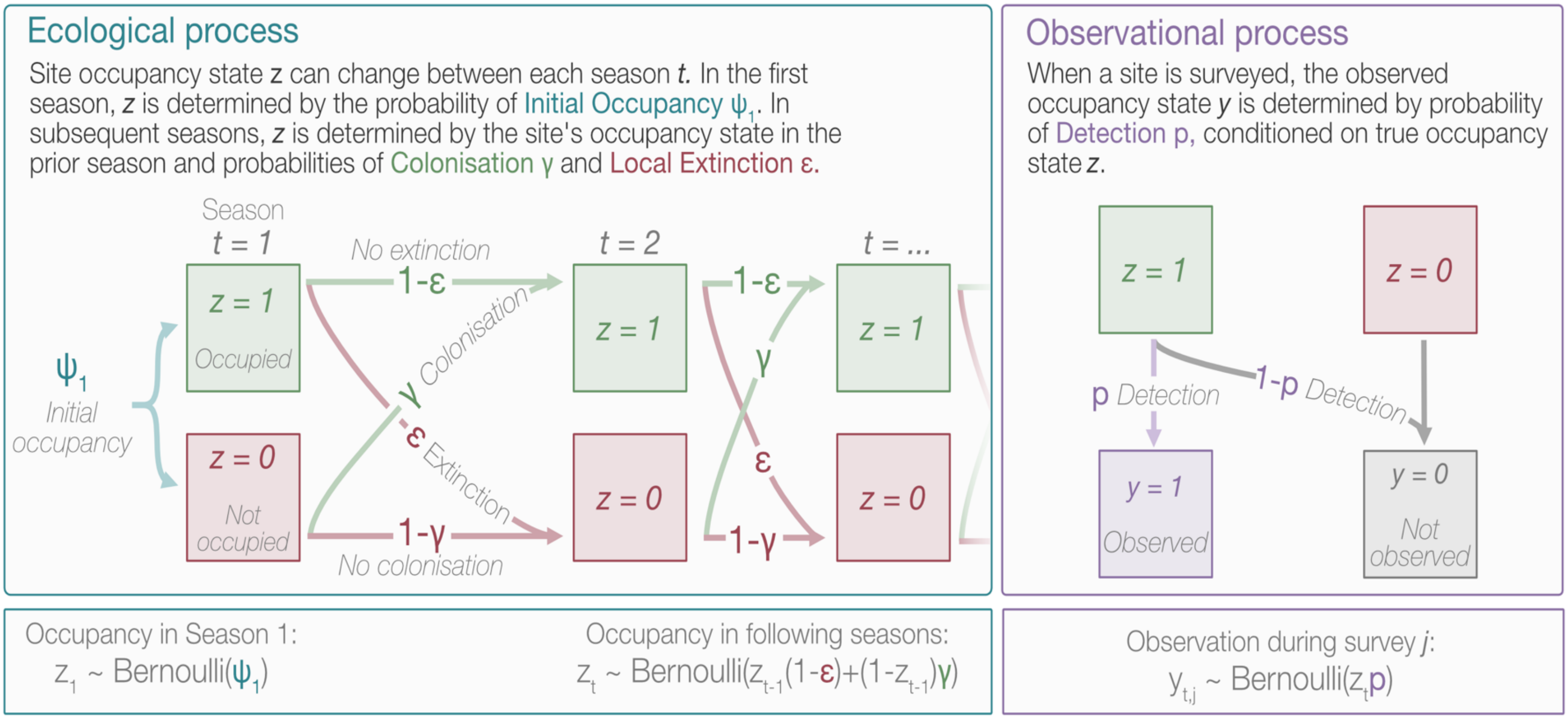

Figure 1: The form of the dynamic occupancy model as described by MacKenzie et al. (2003). The ecological process sub-model describes changes in occupancy over time, while the observational process sub-model describes detectability.

To disentangle the ecological and observational processes the DOM requires a hierarchical sampling design as depicted in Figure 2. Under this design, observations at a site occur during distinct, time-bound seasons within which it is assumed site occupancy does not change (that is, sites are closed to changes in occupancy). In each season multiple observations are made, permitting estimation of the probability of detection conditional on occupancy. Most frequently these repeat observations are collected by revisiting the site on separate occasions, although they can also be attained by alternative means: examples include conducting surveys at multiple locations within a site, using multiple observers during a survey, or recording the time elapsed until a detection is recorded (Guillera-Arroita, 2017; Guillera-Arroita & Lahoz-Monfort, 2017). It is important to note that it is not necessary for the same number of observations to occur in each year or for each site, enabling flexibility in data inputs.

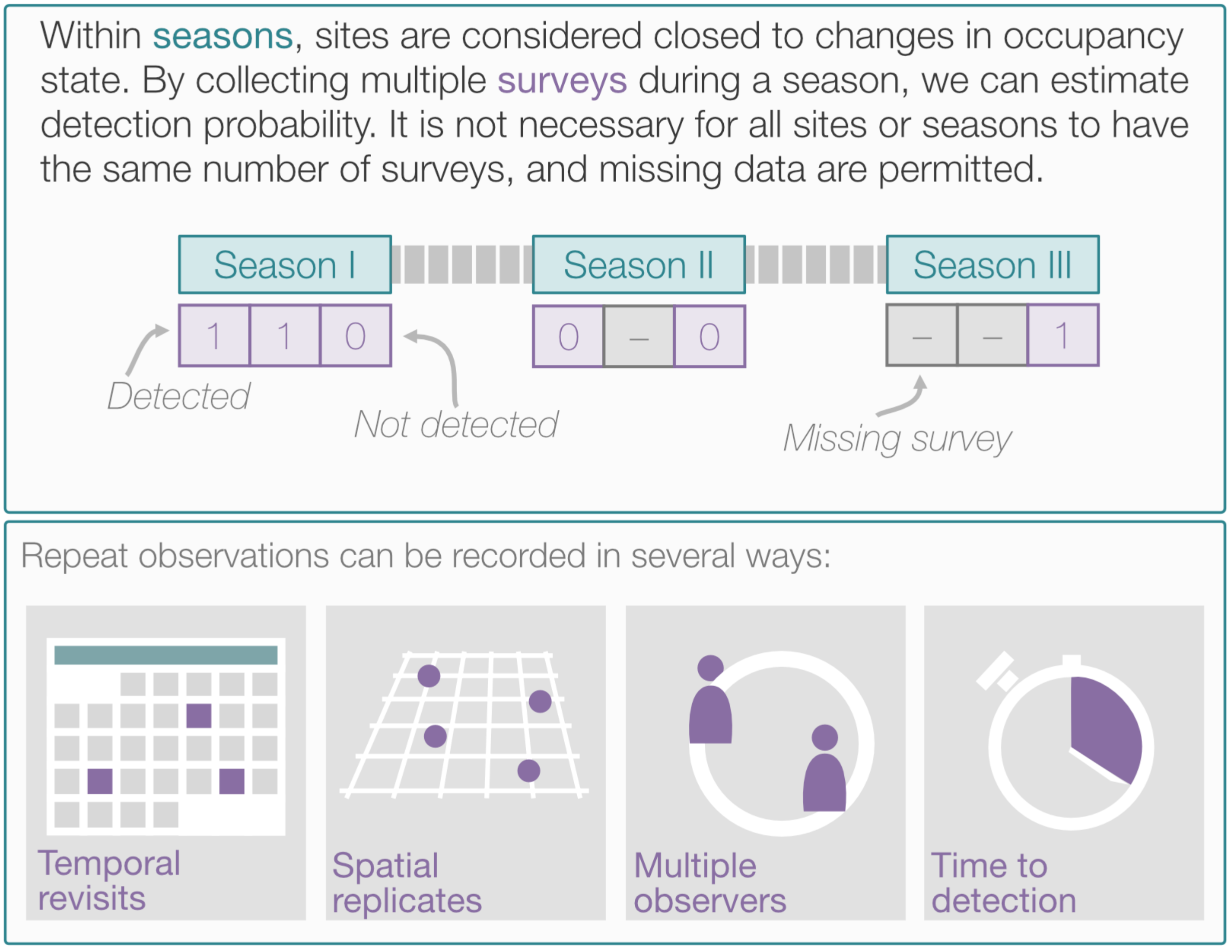

Figure 2: The sampling design of the standard dynamic occupancy model. During seasons (also called primary occasions) sites are considered closed to changes in true occupancy state; occupancy state may only change between seasons. Within each season, multiple observations (also called surveys or secondary occasions) are conducted to record the observed presence or absence of the species at each site. These multiple observations may be recorded in many ways: sites can be revisited multiple times within a season, surveys can be conducted at multiple points within a larger site, multiple observers can conduct surveys simultaneously, or the time

elapsed prior to a detection occurring can be recorded. Note that it is not necessary for each site or season to have the same number of observations, and that missing data can be accommodated.

DOMs make several key assumptions requiring careful consideration by model users; we outline these here as our review interrogates related aspects of model building.

I. False positive detections do not occur. While this assumption can be safely met in many studies, it is not necessarily guaranteed when working with more cryptic species or less reliable survey methods. Previous research has demonstrated the bias induced when false positives occur and are not accounted for, highlighting the need to consider how certain detections truly are for any given study system (McClintock et al., 2010; D. Miller et al., 2015). Importantly, even genuine detections of a species can be considered 'false positives' when they do not represent the intended definition of occupancy, such as detections of transient individuals when the intent is to estimate breeding occupancy (Berigan et al., 2019). Where this assumption cannot reasonably be met, model extensions designed to account for false positive error should be considered (D. Miller et al., 2011; D. Miller et al., 2015; Royle & Link, 2006).

II. Sites are closed to occupancy between seasons. This requirement, best known as the 'closure assumption,' has also been subject to considerable discussion in terms of the bias introduced when it is violated (Otto et al., 2013; Rota et al., 2009). Closure is dependent not only on the life history of the species, but also on the definition of occupancy used by researchers — short seasons may capture dynamics more representative of species 'use,' and it can often be difficult to distinguish local extinction from temporary emigration (Valente et al., 2017). Model extensions to relax the closure assumption have been developed, including approaches using staggered arrival and departure periods between sites (Kendall et al., 2013). A more pertinent approach for most studies, however, is careful consideration of the appropriate definition of occupancy under the survey design used (Steenweg et al., 2018).

III. Heterogeneity in occupancy and detection is accounted for. As with any approach for modelling species occurrence, it is assumed that DOMs appropriately capture variation in occupancy patterns and species detectability across the study system. Generally, this is achieved by allowing model parameters (ψ_1, γ, ε, and ρ) to vary with respect to covariates representing the environmental factors that may be expected to influence these parameters. An important element of this assumption is the caveat that the likelihood of detection of a species can depend not only on the observability of the species, but also on factors like habitat suitability that influence species abundance and activity (Guillera-Arroita, 2017). While no model will fully account for the complexity inherent in patterns of species occupancy and detection, failure to capture key drivers is likely to introduce bias and confound inference made from model estimates (Barry & Elith, 2006). Relative to the first two assumptions, this aspect of DOMs has been less thoroughly discussed and comparatively little is known about how latent heterogeneity can influence model performance.

Since its original description in MacKenzie et al. (2003), the DOM has been further developed with numerous model extensions and alternative formulations including the aforementioned implementations accounting for false positives (D. Miller et al., 2011; D.

Miller et al., 2015; Royle & Link, 2006), as well as models for multiple states beyond occupied and unoccupied (Nichols et al., 2007) and jointly estimated multi-species models (Dorazio et al., 2010). For a comprehensive discussion of common extensions and their applications see Bailey et al. (2014) and Guillera-Arroita (2017), as well as Devarajan et al. (2020) for a more detailed review of multi-species occupancy models.

Systematic review methods

To assess how DOMs have been applied in the years since their introduction we gathered a representative sample of articles that fit DOMs to field ecological data. A pool of candidate articles was generated using two queries on Web of Science. The first of these included all articles from 2004-2023 that cited MacKenzie et al. (2003). To capture any additional relevant articles that did not directly cite MacKenzie et al. (2003), a second query was generated searching articles in the same time-span matching the terms “dynamic occupancy model*”, “multi-season occupancy model*”, OR “occupancy dynamic*”; articles including each of “occupancy”, “colonization”, “extinction|persistence”, AND “detection”; and articles with the term “occupancy” located near “dynamic” in the title, key terms, or abstract. These queries resulted in 1469 articles: 897 were retrieved only from the MacKenzie citations, 274 only from the keywords search, and 298 from both queries. As we were interested in how the use of DOMs has changed through time, we divided all articles across four-year-long strata spanning 2004-2007, 2008-2011, 2012-2015, 2016-2019, and 2020-2024. From each of these strata we randomly selected 20 articles for inclusion in the review. Articles that did not meet inclusion criteria were replaced from within their own strata.

As our review focuses only on applications of the dynamic multi-season occupancy model of MacKenzie et al. (2003) and its extensions, we included articles that fit a model meeting the following criteria: i) used non-simulated, field-collected, presence-absence data; ii) relied on data from multiple sites that could exist in at least two states, including an occupied and unoccupied state; iii) had multiple seasons, between which sites could change states conditional on the prior season's occupancy state and transition probabilities such as colonisation and extinction; and iv) contained at least one parameter describing the detection process.

For each article we recorded key details on authorship, research objectives, study taxa and system, survey methods, and modelling approach. To examine the reasons why authors used DOMs, we allocated each article to one or more categories of objective based on the study's stated aims. These categories were Estimating trends, where authors expressed interest in estimates of site occupancy, colonisation, extinction, or detection probabilities; Testing relationships, where authors explored specific predefined hypotheses of relationships between covariates and model parameters; Identifying drivers, where authors attempted to find which covariates influence model parameters; Predicting temporally, where authors predicted site occupancy under future conditions, Predicting spatially, where authors predicted site occupancy to unsurveyed locations, and Developing methods, where authors introduced, tested, or demonstrated aspects of dynamic occupancy models.

We recorded details on the type of organisms (bird, mammal, etc.) modelled in each article, and how multiple species were modelled where applicable. Taxa were denoted as threatened either when they are listed on the IUCN Red List of Species or when authors indicated that they are threatened. This deference to authors' representation of conservation status was to account for sub-species that lack listings or species that are of more local concern. Study locality and size was documented, the size of the study area being defined as the intended area of inference containing all sites — this was measured to the order of magnitude to account for uncertainty in reporting.

We were particularly interested in how authors navigated the modelling process, from covariate selection through model evaluation. To this end, we recorded all covariates considered in each study regardless of whether they were or were not included in final models. Key traits of each covariate were recorded including their general category, whether they were directly observed or remotely sensed, whether they were static or varied between seasons, and how they were presented in the model (as a linear term, a polynomial term, or as part of an interaction with another covariate (James et al., 2021)). Model selection procedures were also sorted into non-exclusive categories including A priori, where only one model was considered; candidate suite, where a predefined set of models was considered; sequential, where covariates were selected parameter-by-parameter (e.g., fitting detection first, followed by initial occupancy and so on); and simple precursors, where selection was preceded by another simpler model implementation. Finally, for each model we documented whether goodness-of-fit was tested and reported, and whether model performance was assessed by validation with either in or out-of-sample data. For the full spreadsheet of data collected and further details on categorisation, see Supplementary material.

Applications of dynamic occupancy models

A total of 92 articles were included in this review. Based on the acceptance rates within each stratum and the quantity of unprocessed articles, an estimated 496 of the 1152 unreviewed articles in our sample would have met inclusion criteria, with an apparent increase over time in the number of articles fitting DOMs (Figure 3).

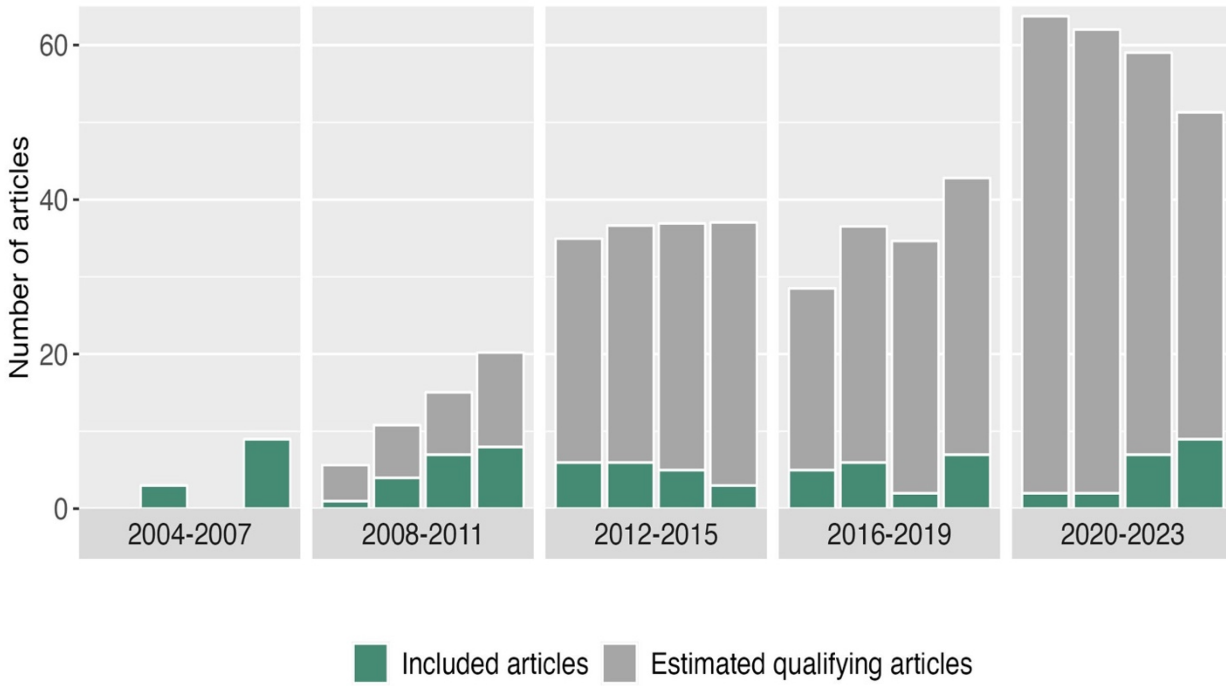

Figure 3: Bars indicate the 92 articles included in our review as a proportion of the estimated number of published articles fitting DOMs, based on the qualification rate for articles in each stratum. The proportion of articles included from each stratum were: 12% from 2005-2008; 24% from 2008-2011; 42% from 2012-2015; 35% from 2016-2019; and 57% from 2020-2023.

The articles included in our review demonstrate considerable diversity in their scope, scale, and objectives. A selection of attributes of these studies is presented in Figure 4, and throughout this section we provide notable examples of DOMs to highlight how this variation appears in practice. Study systems included in our sample are globally distributed: while a majority of articles use data collected in the United States of America, representatives are included from all geographic realms and a broad diversity of ecosystems (Figure 4 A). A particularly notable aspect of these study areas is their variation in size, which ranges from the hyper-local to the continental scale (Figure 4 B). The smallest study locality included in our sample studied insect occurrence in a rainforest plot less than one square kilometre, while the largest analysis modelled avian range shifts across the entire eastern half of the United States (Basset et al., 2023; Clement et al., 2019).

The temporal scale of studies shows similar variability. The time elapsed between the first and last survey ranged from under one month to over forty years, with a median duration of 8 years (Figure 4 F). More meaningful, however, is the period of the primary occasion, as this represents the temporal scale at which changes in occupancy are assumed to occur. Most applications describe primary occasions as occurring annually, although some studies divide years into multiple primary occasions to describe seasonal variation in occupancy. In the most extreme cases the primary occasion may be as brief as a week and capture much finer scale changes in occupancy. These shortened-season occupancy models are most common with camera trapping or acoustic monitoring data, which can be arbitrarily divided into primary and secondary sampling occasions. One example of this is Kleiven et al. (2020)'s study of rodent and mustelid interactions using camera trap data, which used primary occasions less than one week in length. On the other end of the spectrum, some studies modelled primary occasions that were decades apart and represented generational changes in occupancy – this is illustrated in Couturier et al. (2023)'s study on long-term otter recovery in France which used data from two primary occasions, one in 2003 and one in 2012.

DOMs have been used to study a variety of species, although most studies have focused on vertebrate taxa (Figure 4 C). The DOM has been less frequently applied to non-animal organisms, perhaps due to a reduced emphasis on imperfect detection outside of the wildlife modelling community. However, there are exceptions, including DOMs used to model decadal changes in lichen occupancy or the spread of the prion chronic wasting disease (Belinchón et al., 2017; Cook et al., 2022). The application to disease dynamics is not unique, and the DOM has been touted as a valuable tool for such applications (Bailey et al., 2014). Multiple authors have used DOMs to model mosquito dynamics, a strongly applied use-case with important human health implications (Mores et al., 2020; Padilla-Torres et al., 2013). While the vast majority of studies model terrestrial species, there have been a limited number of DOMs fit to data from aquatic systems, including invasive salmon, Great Plains stream fishes, and whales (Falke et al., 2012; Fisher et al., 2014; Pendleton et al., 2022).

Where most studies in our sample fit a model to a single species, 44% fit models to multiple species either as independent models (34%) or explicitly multi-species implementations (12%; some articles attempted multiple approaches). The multi-species models used various extensions of the conventional DOM, including hierarchical models that fit hundreds of species in a single implementation with species-level effects (Dorazio et al., 2010; Hendershot et al., 2020), as well as explicit interaction models that estimate conditional occupancy, colonisation, extinction, and detection probabilities (Fidino et al., 2019; Lesmeister et al., 2015). Rather than using multi-species models, several other authors fit large numbers of independent models to different species (Otto & Roloff, 2012; Peach et al., 2019). Working with large numbers of taxa does raise additional challenges, as the degree to which models are tailored to the taxa is likely to be limited for the sake of practicality.

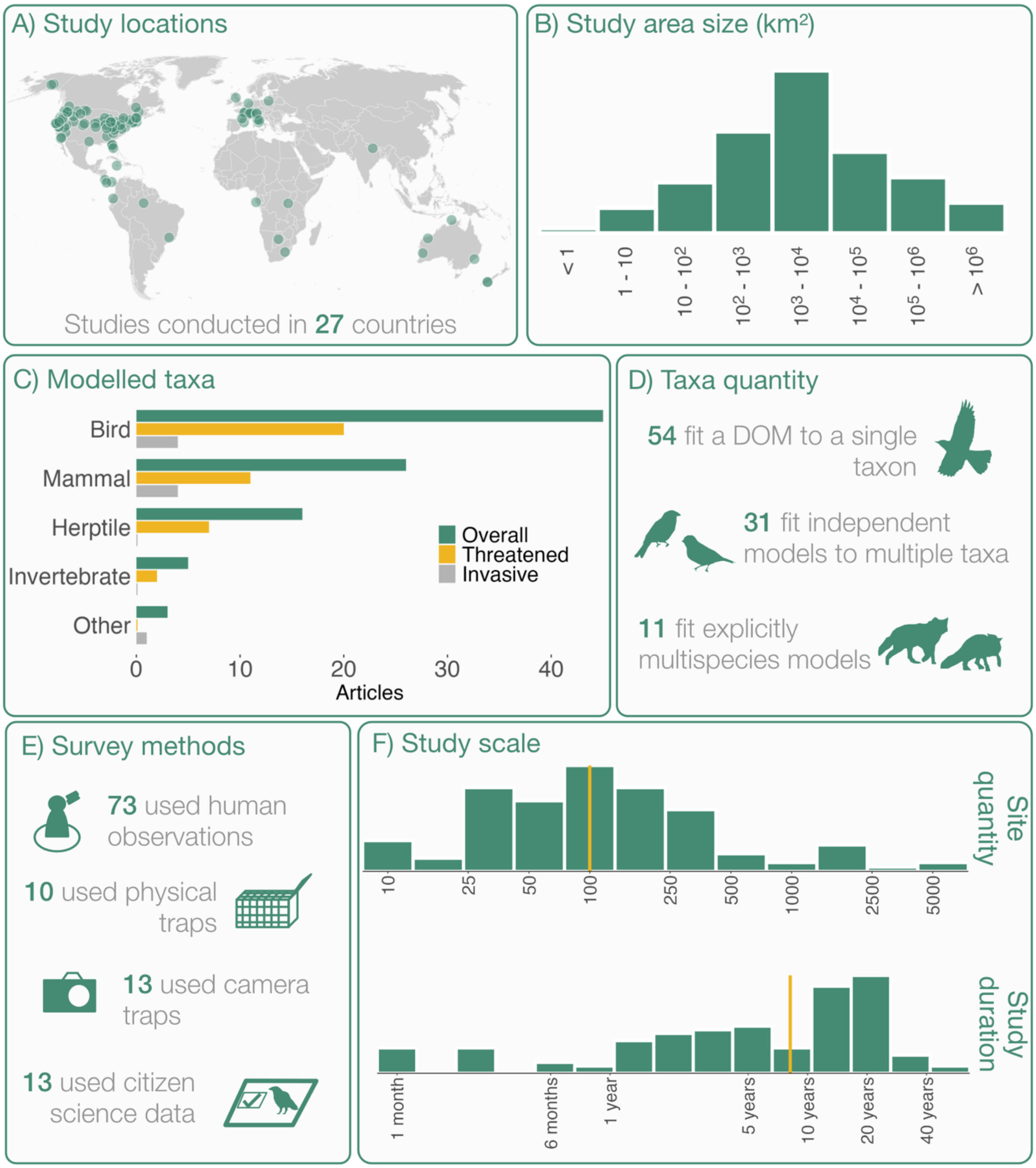

Figure 4: A) Locations of the study areas where data was collected for DOMs. B) Spatial extent of study areas in square kilometres, defined as the area of inference within all surveyed points were contained. C) Number of articles that fit models to

each category of taxa. Taxa were considered 'threatened' if they are listed on the IUCN Red List, or if authors indicate that they are otherwise threatened. D) Explicitly multi-species models include both hierarchical, jointly estimated models as well as more directly interactive models. Some studies fit both independent and multi-species models, such that these values do not sum to our sample size. E) Survey methods used to collect presence/absence data. Note that some articles employed multiple detection methods, and that some methods (e.g., citizen bird counts) may fall into multiple categories. F) Quantity of sites where surveys were conducted and duration of studies. Study duration is defined as the time elapsed between the first and last survey. Yellow lines indicate median values for site quantity (100 sites) and study duration (8.2 years).

Studies collected the presence/absence data required for DOMs from a breadth of different sources and detection methods (Figure 4 E). In our sample, 79% of studies used human observational data, 11% used live-trapping methods, and 14% used detections from camera traps. 10% of articles also use citizen-science data to build detection histories, including both long-term volunteer monitoring programs as well as more ad-hoc surveys of local knowledge (Zuckerberg 2011; Warrier 2020). Using citizen-science data may require additional consideration of assumptions (particularly false-positives), as discussed in greater detail by Cruickshank et al. (2019). Within these broad categories of detection method there is additional diversity, with several studies using unique survey methods to determine the presence of a species at a location. For example, in the only application of a DOM for a marine species in our sample, Pendleton et al. (2022) used aerial transects broken up into grid cells to observe whale occupancy; in another unique implementation, Marescot et al. (2020) fit a unique ’multi-species’ model treating poachers as a taxon and using ranger reports to create detection histories.

Notably, many of these studies use data that were not originally collected in a robust design framework for occupancy modelling. In these articles, authors formatted their data into a hierarchical format post-hoc using a variety of methods. Some defined primary seasons as a discrete time interval, treating all surveys occurring within the season as secondary occasions; others defined sites as larger grid cells, treating any survey falling within the grid as a spatial replicate. These manipulations permit authors to use data that predate the DOM, as in a study using surveys conducted by Joseph Grinnell in 1908 to model century-long changes in occupancy (Riddell et al., 2021).

Additionally, not all articles rely on a single source of detection data; some integrate multiple sources of data to maximise sample size. In one example, Warrier et al. (2020) modelled tiger occupancy in India using three sources of detections: camera traps, sign surveys, and citizen science reports. These integrated detection method models do require additional care and consideration; users must ensure that different detection methods represent comparable spatial and temporal scales, and that any variation in perceptibility between detection methods is accounted for (Pitman et al., 2017). A special case exists when different detection methods are used where one has the potential for false positive detections; e.g., where less-than-certain citizen science detections are combined with certain detections from expert field surveys. In this context, the “certain” detections are used to help account for false-positive detection probability, as in D. Miller et al. (2011)’s study integrating GPS collars and hunter reports to estimate wolf occupancy in Montana.

The flexibility in the data used for DOMs, including the model structure and the scale of observations, is not amenable to a one-size-fits-all definition of occupancy. Users of DOMs must carefully consider precisely what they are modelling and address questions on the scale represented by their model (Chave, 2013). The interpretation of what ‘occupancy’ means and the factors which drive it may differ depending on the scale of what is considered a site, and how that relates to the ecology of a species. This is also true for the temporal scale of occupancy: whether a site is occupied within a week or within a year leads to vastly different conceptions of occupancy. This idea is particularly relevant in cases where the selection of season length is to some extent arbitrary, as with camera-trap or bioacoustic data where continuous recordings can be broken down into distinct 'seasons' of any length. The DOM may be readily applied to these data types, and the proliferation of autonomous survey techniques provides novel opportunities for analyses that are simply not possible with human-collected data (Balantic & Donovan, 2019). This is already apparent in recent camera trap studies that collapsed their detections into seasons of just a few days, far shorter than is realistic with conventional survey methods (Kleiven et al., 2020; Mölle et al., 2022). While these studies provide interesting insights of occupancy at extremely fine temporal scale, further research could help provide general guidance on determining appropriate season and survey durations with respect to research questions and the closure assumption.

Practices in implementation and model building

Building dynamic occupancy models can be a challenging process, requiring careful consideration of which environmental factors to incorporate to adequately represent occupancy and detection in complex natural systems. In this review, we recorded all covariates considered for each model in our sample, including those not used in final models. These are summarised in a taxonomy presented in Table 1, which states the proportion of studies that considered each type of covariate in their models, the means by which that covariate data was collected, and how covariates responses were represented in DOMs. We further delineate their use in the model by classifying covariates into two groups: environmental covariates, representing plausible ecological correlates of parameters; and structural covariates, representing aspects of model form and observation functionally distinct from the environment. Our findings indicate that DOM users have incorporated a wide diversity of covariates in their models — while the most common varieties of covariates include aspects of habitat and land cover, a range of other unique factors were considered by authors in our sample. Many models also incorporate covariates representing aspects of site geometry and connectivity (35% of studies). Often these are included as simple covariates, as in Duggan et al. (2011)'s models using landscape connectivity metrics as covariates on colonisation and extinction. Alternatively, more complex parameterisations explicitly model colonisation or extinction as a spatial process, blending attributes of DOMs and stochastic patch occupancy models (Broms et al., 2016; Risk et al., 2011). Several studies also include biotic interactions with other species as covariates, particularly those focused on taxa threatened by invasive species. This is observed frequently in the Spotted owl literature, where Barred Owl presence is often considered as a covariate driving Spotted owl occupancy dynamics (Olson et al., 2005; Sovern et al., 2014; Dugger et al., 2016). The use of these types of covariates effectively incorporates species interactions in DOMs without requiring the use of more complex explicitly multi-species models.

		Percentage of articles with covariate on parameters:						Percentage which are:		Articles representing this covariate with:	
		Any	ψ1	ψ	γ	ε	ρ	Dynamic	Directly observed	Non-linear relationships	Interactions between covariates
Environmental covariates	**Habitat** *Aspects of habitat and land cover*	55%	41%	25%	43%	46%	28%	24%	33%	12%	25%
	Spatial *Site geometry, connectivity, or other spatial elements*	35%	22%	31%	33%	30%	11%	30%	39%	22%	25%
	Phenology *Time-varying elements distinct from sampling occasions*	33%	1%	0%	5%	4%	33%	100%	0%	41%	9%
	Climate/Weather *Climate, weather, and natural disasters*	33%	13%	12%	18%	18%	24%	77%	35%	32%	26%
	Anthropogenic *Relations to human activity*	25%	20%	6%	23%	21%	6%	11%	8%	12%	20%
	Other environmental *Other environmental covariate not otherwise listed*	21%	5%	19%	4%	10%	13%	71%	78%	0%	0%
	Topography *Elements of landscape topography*	21%	18%	25%	13%	14%	6%	0%	7%	29%	10%
	Biotic interaction *Interactions with other (non-plant) species*	15%	7%	0%	14%	14%	5%	59%	64%	7%	20%
	Hydrology *Aspects of hydrology, such as distance to water*	14%	8%	25%	14%	11%	4%	37%	33%	29%	21%
	Any Environmental	***91%***	***62%***	***94%***	***74%***	***73%***	***70%***	***43%***	***30%***	***33%***	***26%***
Structural covariates	**Primary occasion** *Effect of the primary occasion*	65%	1%	44%	39%	38%	61%	99%	0%	15%	8%
	Observation *Details on the observation process*	24%	0%	0%	0%	0%	24%	93%	7%	8%	4%
	Secondary occasion *Effect of the secondary occasion*	15%	0%	0%	0%	0%	15%	100%	10%	13%	0%
	Site effect *Site-level effects*	3%	0%	0%	2%	2%	2%	0%	0%	0%	0%
	Other structural *Other structural covariate not otherwise listed*	3%	1%	0%	0%	0%	2%	33%	33%	0%	33%
	Species effect *Species-level effects*	2%	2%	0%	2%	2%	1%	0%	0%	0%	50%
	Any Structural	***81%***	***3%***	***44%***	***41%***	***40%***	***80%***	***94%***	***4%***	***16%***	***8%***
All covariates	**All covariates**	**99%**	**63%**	**100%**	**85%**	**85%**	**97%**	**54%**	**20%**	**35%**	**24%**

Table 1: All covariates considered for inclusion in a study were classified into mutually exclusive categories. We calculate the percentage of studies that include at least one covariate of a given category on any parameter, Initial Occupancy ($\psi 1$), Occupancy (ψ), Colonisation (γ), Extinction(ε), and Detection (ρ). We also present the average percentage of covariates in a study that are dynamic (varying through seasons) and directly observed for each category, as well as the percentage of studies that model each category of covariate with a non-linear relationship or as part of an interaction with another covariate.

Covariate data for studies in our sample was either collected directly by researchers (an average of 30% of environmental covariates per study), or derived from pre-existing remotely sensed datasets (70% of environmental covariates); this of course varies depending on the category of covariate (Table 1). Directly collected data can often represent finer-scale factors like prey species occurrence or details of habitat structure, which can be difficult to measure remotely but can often be more proximal drivers of occupancy. These species-specific covariates do come with trade-offs, as they can be expensive to collect and often preclude projection to locations where these data are unavailable. For studies interested in making such projections, remotely-sensed covariates are generally more feasible despite their generally more distal nature (Austin, 2002). An average of 43% percent of environmental factors and 94% of structural factors included in reviewed studies were dynamic covariates that varied through time — this again varied with the category of covariate in question, with terms relating to climate or weather most frequently dynamic and topographic covariates universally static (Table 1).

In the standard DOM, covariates for each parameter are most commonly incorporated via a logistic regression (i.e., a linear regression through a logit link function) (MacKenzie et al., 2017). Statistical relationships between model parameters and covariates (e.g., between initial occupancy and its environmental covariates) are represented as linear terms unless more complexity is specified. Of course, many ecological relationships are non-linear and require more complex forms to be realistically represented in a model. Austin (2007) discusses the importance of modelling ecologically realistic responses to covariates, advocating for careful consideration of the most appropriate statistical form for hypothesised relationships. Non-linear responses can be easily accommodated in DOMs by using polynomial transformations and interactions between covariates. Despite this, in our sample only 35% of articles tested one or more non-linear responses to an environmental covariate, with the majority of studies representing all covariates as simple linear terms (Table 1). Interactions between covariates were similarly rare, with only 24% of studies considering at least one interaction between terms (Table 1). The relatively low emphasis on more complex non-linear responses contrasts with other popular methods for modelling species occupancy. 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). This emphasis on more complex responses in SDMs may be due to their frequent application across relatively large geographic extents that might encompass the full species niche, where environmental relationships may be expected to be non-linear. However, as previously indicated DOMs have also been implemented across large spatial extents where the same assumptions should exist and similar responses might be expected. Similar concerns exist for the relatively low level of covariate interactions, given that these relationships are commonly expected based on ecological theory and that their exclusion can negatively impact model performance (Guisan et al., 2006).

In addition to the types of covariates considered for modelling we also recorded the size of the candidate covariate pool in each reviewed study, tallying the number of environmental and structural covariates that were available for use on each parameter (Figure 5). This is an area of considerable variation in modelling practices — the number of covariates considered for each parameter ranges from 0 (effectively modelling the parameter as a constant) to over 40 candidates on a single parameter.

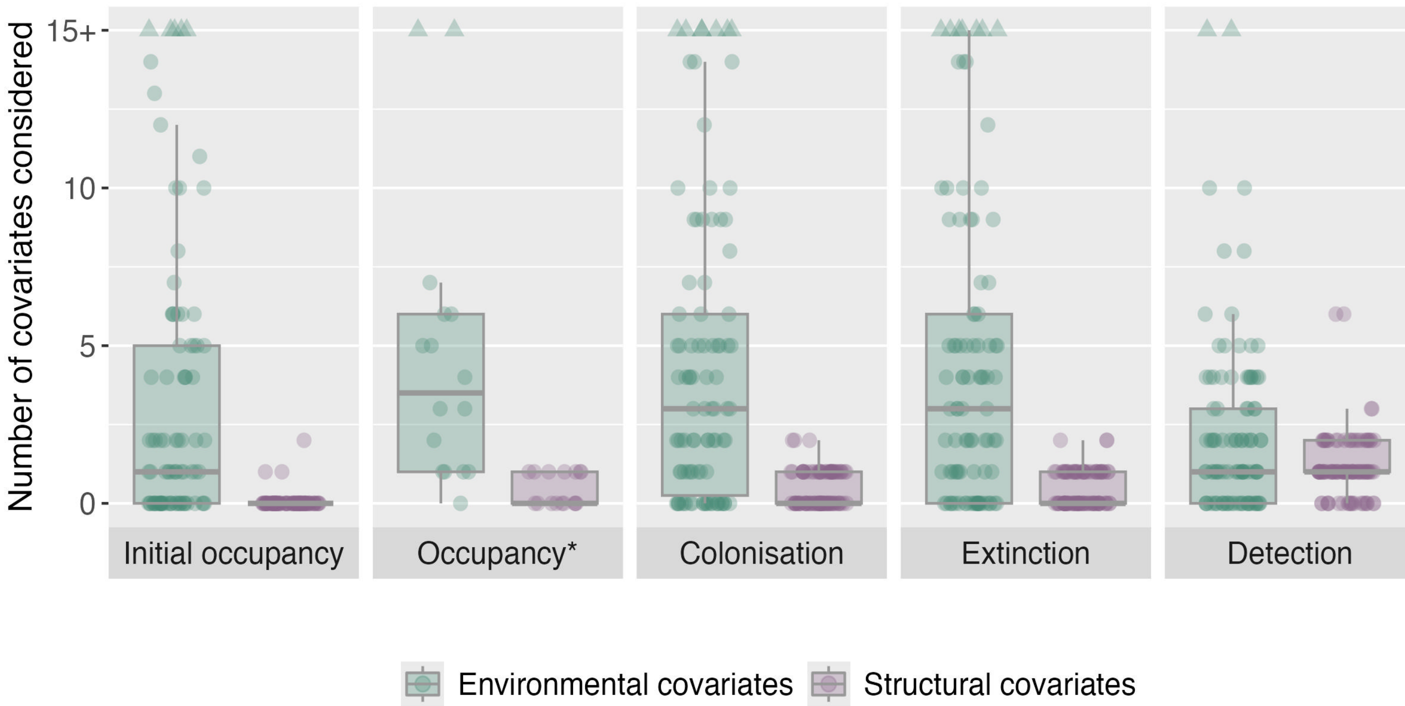

Figure 5: The number of covariates considered for each parameter across all studies in our sample. 'Occupancy' given here represents the alternative parameterisation of the DOM that jointly estimates Occupancy for every season, Colonisation, and Detection, where Extinction is a derived parameter; this differs from the more popular Initial occupancy/Colonisation/Extinction/Detection parameterisation. Here, a 'covariate' is defined as each distinct covariate considered for inclusion. Linear and quadratic representations of the same covariate are counted as one covariate.

The median number of covariates varies strongly by parameter, with transition probabilities (colonisation and extinction) more likely to consider a broader range of environmental covariates compared to initial occupancy and detection. The lack of any covariates considered for initial occupancy in many cases (38% of studies) is particularly notable — unless a study is conducted at very small extents or study sites are truly uniform in their suitability, one would expect some amount of non-random variation in occupancy probability across any study system that will not be captured when representing initial occupancy as a constant. Omission of the factors that drive occupancy is a known source of bias in static SDMs (Barry & Elith, 2006), which the initial occupancy component of the DOM conceptually resembles. Furthermore, any bias in occupancy estimation in the first time-step will perpetuate into future seasons due to the DOM's Markovian nature, with implications for the reliability of model outputs. Detection probability is also typically represented with low numbers of environmental covariates (Table 1). Recall that detectability is dependent on not only observation factors, but also by drivers of abundance and species use; the low number of candidate covariates here relative to colonisation and extinction raises additional questions on whether this variation is always appropriately captured in models.

To this point in our review, we have discussed only the covariates that were considered for inclusion in DOMs without regard for which terms were actually included in the final models used by authors to make their inference. Identifying the 'best' model from a range of possible candidates can be a challenging process, especially for hierarchical models like DOMs (Doherty et al., 2012). Consider that the quantity of candidate models involving up to n covariates on p parameters can be given by $(2^{n})^{p}$: where a simple linear regression may have up to 64 possible combinations of 6 covariates, a 4-parameter dynamic occupancy model would have 16 million combinations of these same covariates. This explosion of candidate models generally precludes exhaustive comparisons of possible models and requires some reduction of the models tested. Models in our sample include both Frequentist (n = 76) and Bayesian (n = 24) frameworks that differ considerably in their manner of implementation. A summary of modelling practices in DOMs is given in Table 2. Note that some articles in our sample included multiple distinct modelling workflows, resulting in a higher total number of models than articles.

Table 2: Modelling practices in dynamic occupancy models, subset by frequentist or Bayesian implementations. The median covariate count presented here represents the median quantity of covariates considered for each model parameter across the studies in our review. The model selection methods represented in this table are non-exclusive and some articles employ multiple approaches. 2 models included in the 'Overall' column are neural network based and fall into neither the Frequentist or Bayesian categories.

	Frequentist	Bayesian	All models
Number of studies	76	24	*102*
Median covariates considered per parameter	3	2.12	*2.75*
Covariate selection methods			
Percentage using any model selection approach	95%	33%	*80%*
Percentage comparing models in a candidate set	45%	12%	*36%*
Percentage using procedural model selection	37%	0%	*27%*
Percentage selecting covariates with simpler models	8%	4%	*7%*
Percentage using model-averaging	47%	4%	*36%*
Model evaluation conducted			
Percentage calculating goodness-of-fit	20%	12%	*18%*
Percentage assessing predictive performance	4%	17%	*7%*

The largest differences between the frequentist and Bayesian models in our sample lie in their approaches to model selection. Where 95% of frequentist models perform some manner of model selection to determine covariate inclusions for their models, only 33% of Bayesian models do so, with the majority instead fitting a single model defined a priori. Methods used for model selection vary considerably both within and between the two modelling frameworks (Table 2). For frequentist models, the most popular and conventional approach to model selection (45% of models) involves the creation of a fully pre-defined model set containing some number of hypothesis candidate models, where the best model is selected according to the lowest AIC score. The next most popular method in frequentist studies is to use procedural model selection methods (37% of models), where the structure for each model parameter is fit in sequence. For example, this protocol might first identify the best structure for detection probability while holding the other parameters constant, before moving on to initial occupancy and so on until all parameters are fixed. The remainder of frequentist studies (8%) use a variety of approaches, such as fitting simpler models like single season occupancy models to identify the most informative terms to use in a DOM. Across all frequentist implementations, 47% of articles summarise a final subset of models using multi-model inference and model-averaging with AIC weights (Burnham & Anderson, 2004).

Those Bayesian models which do perform model selection take various approaches, with largely idiosyncratic methods across these studies. While direct comparison of model fit is rare in Bayesian methods, it is feasible — Urban et al. (2023) identifies the best model from a Bayesian candidate set using the predictive performance on both in and out-of-sample validation data. Another approach used by Cook et al. (2022) fit a global model including all covariates before removing each covariate where the 95% credible interval of the posterior distribution overlapped zero and refitting the model. Ahumada et al. (2013) takes a hybrid approach, in which model selection is conducted by a procedural method in the frequentist framework before refitting the best structure as a Bayesian model.

Limited research has been conducted on the advantages of different methods for covariate selection in DOMs, and there is unlikely to be a one-size-fits-all approach that will be appropriate for all possible use-cases. However, it is important to consider the implications of the different model selection approaches in common usage, and research from the SDM literature highlights how covariate selection can influence our interpretation of model outputs (Brodie et al., 2020). Within the occupancy modelling literature, Stewart et al. (2023)'s article on covariate selection in single-season occupancy models discusses important attributes of the information-criteria approaches widely used in frequentist models in our review, noting that these approaches can lead to inaccurate coefficient estimates that may influence model inference. Morin et al. (2020)'s work raises other concerns on procedural model selection methods, demonstrating how the fine details of modelling protocols can determine whether the best-fitting possible model is identified and which covariates appear in final models. The literature on performance of Bayesian model selection methods is more sparse, although Hooten & Hobbs (2015)'s guide to Bayesian model selection in ecology is a valuable resource for possible methods of fitting those models.

Regardless of covariate selection protocol, ‘the selection of a best model does not guarantee the selection of a good model’ (MacKenzie & Bailey, 2004). Assessment of the performance of DOMs generally requires additional steps, and the best ways to achieve this are not yet clear. There is no broadly-accepted goodness-of-fit test for dynamic occupancy models, although MacKenzie & Bailey (2004)’s approach for single-season occupancy models using a parametric bootstrap that can been extended to DOMs; this test is implemented in the AICcModAvg and unmarked R packages (Mazerolle 2016; Kéry and Chandler, 2016). Kéry & Royle (2021) describe the test and present an alternative based on separately assessing fit to static and dynamic components of the model. In Bayesian implementations, posterior predictive checks offer means to assess model fit (Gelman, 2014). Broms et al. (2016) discusses further possibilities for model evaluation in the Bayesian context focusing on single-season multi-species occupancy models, and extensions of their approach may also be applicable to the DOM. As with other hierarchical models, model evaluation for DOMs can be difficult and somewhat uncertain compared to other model types, as the primary response variable of interest (species occupancy) is a latent variable where the true state is generally not known. Predictive performance evaluation is thus typically based on observed occupancy data, where a DOM is used to simulate detection histories to be compared with field survey results.

Perhaps as a result of these difficulties, assessment of model fit and model performance was rare amongst the articles in our sample. Only 18% of studies tested for goodness-of-fit, and just 7% calculated predictive performance. These rates are considerably lower than those reported for SDMs; for which closer to 50% of articles were found to test both fit and performances (Araújo et al., 2019).

Modelling objectives

Studies in our sample were classified according to their objectives as stated by authors to explore the use-cases that the DOM has been applied to (Figure 6). These objective categories are non-exclusive, with many studies falling into multiple categories. Across the study period the DOM has been frequently employed to estimate trends in species occupancy, to explore relationships between environmental factors and occupancy, and, increasingly often, to make predictions of occupancy spatially to unsurveyed locations or temporally to hypothetical future conditions. Within each of these broad categories lies even more variation in objectives, emphasising the DOM's flexibility as a tool for making ecological inference in diverse contexts.

36% of articles used the DOM to monitor trends in occupancy state through time ("Estimate trends" in Figure 6 A), both for single species of high conservation interest and for broader community assemblages of species (Ahumada et al., 2013; Scott & Rissler, 2015). These studies offer valuable insights on the state of species across landscapes and through time, demonstrating the DOM's suitability for monitoring oriented projects; including those where multiple species need to be assessed at once.

Other articles were more focused on examining the factors that influence species occupancy, either testing pre-specified hypothetical relationships (44% of articles) or taking a broader tack to identify drivers of occupancy without preconception (35%). The distinction between the two is important, as it guides which covariates may be considered and how model selection may be used to fit models. Many of these studies target core conservation priorities for their focal species, like Olson et al. (2005)'s early DOM assessing the influence of barred owls on threatened spotted owl. Explorations of these pivotal relationships are important for guiding management action, and may also be used to test the effectiveness of these actions (K. E. Miller & Brown, 2023). In scenarios where less is known about species habitat preference, DOMs may be used to examine the influence of a wider variety of factors as in Huber et al. (2017)'s study testing the relative influence of dozens of habitat covariates on Wood warbler occupancy. In recent years, increasing numbers of articles have used DOMs to generate predictions of species occupancy; this is a use-case for which DOMs show considerable promise (Briscoe et al., 2021; Kéry et al., 2013). These predictions often have strong utility for conservation management. For example, McGowan et al. (2020) projects future occupancy for wetland birds under alternative management scenarios, and Pollentier et al. (2021) generates distribution maps resembling those made with SDMs.

Figure 6: A) Proportion of articles in each year-strata and across all years that match each of six non-exclusive objective categories. B) Quantity of covariates considered per parameter for models that pursued each objective.

Study objectives strongly inform approaches to model building, particularly with respect to covariate selection. Figure 6 B presents the quantity of covariates considered for models applied to each of our objective categories, with apparent differences between objectives. Not unexpectedly, articles that focused on describing new methods for DOMs had fewer covariates than those studies focused on more applied objectives. However, differences persist between those articles observing trends, identifying trends in occupancy, testing relationships, and making predictions.

Synthesis and key priorities

Approaches to building any type of model will necessarily depend on the possibilities of the data at hand and on the priorities of the model-builder. This precludes any prescription of the 'best' way to build a model, however, there are still important discussions to be had on decisions made in the modelling process. One aspect of fitting DOMs meriting broader discussion centres on 'model complexity,' and how much must be incorporated into models to reliably model species occupancy under different contexts and use cases. Complexity is a broad term encompassing many aspects of a model (Merow et al., 2014), and opinions on simplicity versus complexity in ecological models are diverse. Where some advocate for the simplest possible models, arguing that they are most generalisable; others insist that overly-simple models cannot adequately represent the most important drivers in a system (Evans et al., 2013; Lonergan, 2014). By their nature the DOM is somewhat more complex than simpler models for studying occupancy due to their hierarchical structure, which is necessary to control for detectability and to capture occupancy dynamics. Within this structure, however, further complexity is to some degree up to the modeller: one can choose how many covariates to consider for inclusion on the various parameters, and how to represent the nature of the relationship between those covariates and parameters. Research from SDMs indicates that allowing for more complex relationships can improve model performance in predicting occupancy (Valavi et al., 2023), an increasingly popular use-case for the DOM. Within the DOM framework, there are promising developments on that front: Joseph (2020) presents a novel neural-network occupancy model that allows for exponentially higher levels of complexity and may offer improved performance for prediction-oriented studies.

Covariate selection seems to be a particularly important area for further investigations into building DOMs. In our review, we see little consensus around which approaches are most applicable for any given use case, and existing work on covariate selection in DOMs raises concerns on whether common methods always produce the most suitable models (Morin et al., 2020; Stewart et al., 2023). This is true for both frequentist and Bayesian implementations, and comparative research on covariate selection under both frameworks may help to inform model users on which method is most appropriate for their use-cases. In a similar vein, the low number of articles in our review that conducted model evaluation or assessed model fit raises different questions. While the appropriate method of model evaluation may depend on data availability and research objectives, assessing models by some method is generally important to understand how reliable model outputs may be (Araújo et al., 2005;

Guisan & Thuiller, 2005). Existing uncertainties around whether current methods are suitable for the task likely discourage users from calculating these metrics, and further research is needed to establish trusted practices for assessing the quality of DOMs.

These priorities are particularly important given the frequent applied objectives of DOM users, tackling challenges that include assessing critically endangered species, guiding public health management of disease vectors, and tracking rapidly developing biological invasions (Carvalho et al., 2023; Moreira et al., 2016; Wood et al., 2020). The DOM is well suited to these situations, and it is to be expected that as these types of applications are more commonly attempted understanding the sensitivity of model outputs to decisions made in the model fitting process becomes increasingly important. In the two decades since the publication of MacKenzie et al. (2003) the dynamic occupancy model has become a widely used tool for ecological inference, with numerous extensions to the modelling framework further broadening the scope of questions and use-cases for which it may be applied. Given their increasing popularity, further research and guidelines around issues of model building may help to make DOMs more accessible to newcomers and support confidence in model interpretation. Parallels to existing work in the SDM literature on guidelines for reporting and modelling standards could be valuable contributions in achieving these aims and provide useful examples in developing methods and guidelines specifically focused on DOMs (Araújo et al., 2019; Zurell et al., 2020).

References

Ahumada, J. A., Hurtado, J., & Lizcano, D. (2013). Monitoring the Status and Trends of Tropical Forest Terrestrial Vertebrate Communities from Camera Trap Data: A Tool for Conservation. *PLOS ONE*, *8*(9), e73707. https://doi.org/10.1371/journal.pone.0073707

Araújo, M., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, *5*(1), eaat4858. https://doi.org/10.1126/sciadv.aat4858

Araújo, M., Pearson, R., Thuiller, W., & Erhard, M. (2005). Validation of species-climate impact models under climate change. *Global Change Biology*, *11*, 1504–1513. https://doi.org/10.1111/j.1365-2486.2005.01000.x

Austin, M. (2002). Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling*, *157*(2), 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling*, *200*(1), 1–19. https://doi.org/10.1016/j.ecolmodel.2006.07.005

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

Balantic, C., & Donovan, T. (2019). Dynamic wildlife occupancy models using automated acoustic monitoring data. *Ecological Applications*, *29*(3). https://doi.org/10.1002/eap.1854

Barry, S., & Elith, J. (2006). Error and uncertainty in habitat models. *Journal of Applied Ecology*, *43*(3), 413–423. https://doi.org/10.1111/j.1365-2664.2006.01136.x

Basset, Y., Butterill, P. T., Donoso, D. A., P. A. Lamarre, G., Souto-Vilarós, D., Perez, F., Bobadilla, R., Lopez, Y., Alejandro Ramírez Silva, J., & Barrios, H. (2023). Abundance, occurrence and time series: Long-term monitoring of social insects in a tropical rainforest. *Ecological Indicators*, *150*, 110243. https://doi.org/10.1016/j.ecolind.2023.110243

Belinchón, R., Harrison, P. J., Mair, L., Várkonyi, G., & Snäll, T. (2017). Local epiphyte establishment and future metapopulation dynamics in landscapes with different spatiotemporal properties. *Ecology*, *98*(3), 741–750. https://doi.org/10.1002/ecy.1686

Berigan, W. J., Jones, G. M., Whitmore, S. A., Gutiérrez, R. J., & Peery, M. Z. (2019). Cryptic wide-ranging movements lead to upwardly biased occupancy in a territorial species. *Journal of Applied Ecology*, *56*(2), 470–480. https://doi.org/10.1111/1365-2664.13265

Bertelsmeier, C., Luque, G. M., & Courchamp, F. (2013). Increase in Quantity and Quality of Suitable Areas for Invasive Species as Climate Changes. *Conservation Biology*, *27*(6), 1458–1467. https://doi.org/10.1111/cobi.12093

Briscoe, N. J., Elith, J., Salguero-Gómez, R., Lahoz-Monfort, J. J., Camac, J. S., Giljohann, K. M., Holden, M. H., Hradsky, B. A., Kearney, M. R., McMahon, S. M., Phillips, B. L., Regan, T. J., Rhodes, J. R., Vesk, P. A., Wintle, B. A., Yen, J. D. L., & Guillera-Arroita, G. (2019). Forecasting species range dynamics with process-explicit models: matching methods to applications. *Ecology Letters*, *22*(11), 1940–1956. https://doi.org/10.1111/ele.13348

Briscoe, N. J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A.-K., Kéry, M., Schmid, H., & Guillera-Arroita, G. (2021). Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. *Global Change Biology*, *27*(18), 4269–4282. https://doi.org/10.1111/gcb.15723

Brodie, S. J., Thorson, J. T., Carroll, G., Hazen, E. L., Bograd, S., Haltuch, M. A., Holsman, K. K., Kotwicki, S., Samhouri, J. F., Willis-Norton, E., & Selden, R. L. (2020). Trade-offs in covariate selection for species distribution models: a methodological comparison. *Ecography*, *43*(1), 11–24. https://doi.org/10.1111/ecog.04707

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

Burnham, K. P., & Anderson, D. R. (2004). *Model Selection and Multimodel Inference*. Springer. https://doi.org/10.1007/b97636

Carvalho, E. a. R., Mendonça, E. N., Lopes, A. M. C., & Haugaasen, T. (2023). Current status of the Critically Endangered Black-winged Trumpeter Psophia obscura in one of its last strongholds. *Bird Conservation International*, *33*, e12. https://doi.org/10.1017/S0959270922000077

Chave, J. (2013). The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters*, *16*(s1), 4–16. https://doi.org/10.1111/ele.12048

Clement, M. J., Nichols, J. D., Collazo, J. A., Terando, A. J., Hines, J. E., & Williams, S. G. (2019). Partitioning global change: Assessing the relative importance of changes in climate and land cover for changes in avian distribution. *Ecology and Evolution*, *9*(4), 1985–2003. https://doi.org/10.1002/ece3.4890

Cook, J. D., Williams, D. M., Porter, W. F., & Christensen, S. A. (2022). Improved predictions and forecasts of chronic wasting disease occurrence using multiple mechanism dynamic occupancy modeling. *The Journal of Wildlife Management*, *86*(7), e22296. https://doi.org/10.1002/jwmg.22296

Couturier, T., Steinmetz, J., Defos du Rau, P., Marc, D., Trichet, E., Gomes, R., & Besnard, A. (2023). Intensive agriculture as the main limiting factor of the otter's return in southwest france. *Biological Conservation*, *279*, 109927. https://doi.org/10.1016/j.biocon.2023.109927

Cruickshank, S. S., Bühler, C., & Schmidt, B. R. (2019). Quantifying data quality in a citizen science monitoring program: False negatives, false positives and occupancy trends. *Conservation Science and Practice*, *1*(7). https://doi.org/10.1111/csp2.54

Devarajan, K., Morelli, T. L., & Tenan, S. (2020). Multi-species occupancy models: review, roadmap, and recommendations. *Ecography*, *43*(11), 1612–1624. https://doi.org/10.1111/ecog.04957

Doherty, P. F., White, G. C., & Burnham, K. P. (2012). Comparison of model building and selection strategies. *Journal of Ornithology*, *152*(2), 317–323. https://doi.org/10.1007/s10336-010-0598-5

Dorazio, R. M., Kéry, M., Royle, J. A., & Plattner, M. (2010). Models for inference in dynamic metacommunity systems. *Ecology*, *91*(8), 2466–2475. https://doi.org/10.1890/09-1033.1

Dormann, C. F. (2007). Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, *8*(5), 387–397. https://doi.org/10.1016/j.baae.2006.11.001

Duggan, J. M., Schooley, R. L., & Heske, E. J. (2011). Modeling occupancy dynamics of a rare species, Franklin's ground squirrel, with limited data: are simple connectivity metrics adequate? *Landscape Ecology*, *26*(10), 1477–1490. https://doi.org/10.1007/s10980-011-9652-9

Dugger, K. M., Forsman, E. D., Franklin, A. B., Davis, R. J., White, G. C., Schwarz, C. J., Burnham, K. P., Nichols, J. D., Hines, J. E., Yackulic, C. B., Doherty, P. F. ,. Jr., Bailey, L., Clark, D. A., Ackers, S. H., Andrews, L. S., Augustine, B., Biswell, B. L., Blakesley, J., Carlson, P. C., … Sovern, S. G. (2016). The effects of habitat, climate, and barred owls on long-term demography of northern spotted owls. *The Condor*, *118*(1), 57–116. https://doi.org/10.1650/CONDOR-15-24.1

Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, *1*(4), 330–342. https://doi.org/10.1111/j.2041-210X.2010.00036.x

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., Langhe, R. de, Lessells, C. M., Merz, M., O'Malley, M. A., Orzack, S. H., Weisberg, M., Wilkinson, D. J., Wolkenhauer, O., & Benton, T. G. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution*, *28*(10), 578–583. https://doi.org/10.1016/j.tree.2013.05.022

Falke, J. A., Bailey, L. L., Fausch, K. D., & Bestgen, K. R. (2012). Colonization and extinction in dynamic habitats: an occupancy approach for a Great Plains stream fish assemblage. *Ecology*, *93*(4), 858–867. https://doi.org/10.1890/11-1515.1

Fidino, M., Simonis, J. L., & Magle, S. B. (2019). A multistate dynamic occupancy model to estimate local colonizationextinction rates and patterns of co-occurrence between two or more interacting species. *Methods in Ecology and Evolution*, *10*(2), 233–244. https://doi.org/10.1111/2041-210X.13117

Fisher, A. C., Volpe, J. P., & Fisher, J. T. (2014). Occupancy dynamics of escaped farmed Atlantic salmon in Canadian Pacific coastal salmon streams: implications for sustained invasions. *Biological Invasions*, *16*(10), 2137–2146. https://doi.org/10.1007/s10530-014-0653-x

Franklin, J. (2010). *Mapping species distributions: Spatial inference and prediction*. Cambridge University Press. https://www.cambridge.org/core/books/mapping-species-distributions/58225AE5693AED8BD812F7CEBE35378A

Gelman, A. (2014). *Bayesian data analysis*. https://research.ebsco.com/linkprocessor/plink?id=86df87c2-efc9-3fae-8c85-32142daec7af

Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlifehabitat models. *Biological Conservation*, *116*(2), 195–203. https://doi.org/10.1016/S0006-3207(03)00190-3

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, *40*(2), 281–295. https://doi.org/10.1111/ecog.02445

Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2017). Species occupancy estimation and imperfect detection: shall surveys continue after the first detection? *AStA Advances in Statistical Analysis*, *101*(4), 381–398. https://doi.org/10.1007/s10182-017-0292-5

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. Mc. C., Aspinall, R., & Hastie, T. (2006). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, *43*(3), 386–392. https://doi.org/10.1111/j.1365-2664.2006.01164.x

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, *8*(9), 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Gutiérrez-Arellano, C., Crone, E. E., Pettorelli, N., & Hodgson, J. A. (2024). Broadening applications of stochastic patch occupancy models over three decades. *Diversity and Distributions*, *n/a*(n/a), e13822. https://doi.org/10.1111/ddi.13822

Hendershot, J. N., Smith, J. R., Anderson, C. B., Letten, A. D., Frishkoff, L. O., Zook, J. R., Fukami, T., & Daily, G. C. (2020). Intensive farming drives long-term shifts in avian community composition. *Nature*, *579*(7799), 393–396. https://doi.org/10.1038/s41586-020-2090-6

Hooten, M. B., & Hobbs, N. T. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs*, *85*(1), 3–28. https://doi.org/10.1890/14-0661.1

Huber, N., Kéry, M., & Pasinelli, G. (2017). Occupancy dynamics of the Wood Warbler Phylloscopus sibilatrix assessed with habitat and remote sensing data. *Ibis*, *159*(3), 623–637. https://doi.org/10.1111/ibi.12472

Humboldt, A. von. (1849). *Cosmos : a sketch of a physical description of the universe*.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer US. https://doi.org/10.1007/978-1-0716-1418-1

Joseph, M. B. (2020). Neural hierarchical models of ecological populations. *Ecology Letters*, *23*(4), 734–747. https://doi.org/10.1111/ele.13462

Kellner, K. F., & Swihart, R. K. (2014). Accounting for Imperfect Detection in Ecology: A Quantitative Review. *PLOS ONE*, *9*(10), e111436. https://doi.org/10.1371/journal.pone.0111436

Kendall, W. L., Hines, J. E., Nichols, J. D., & Grant, E. H. C. (2013). Relaxing the closure assumption in occupancy models: staggered arrival and departure times. *Ecology*, *94*(3), 610–617. https://doi.org/10.1890/12-1720.1

Kéry, M., Guillera-Arroita, G., & Lahoz-Monfort, J. J. (2013). Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, *40*(8), 1463–1474. https://doi.org/10.1111/jbi.12087

Kéry, M., & Royle, J. A. (2021). *Applied hierarchical modeling in ecology : analysis of distribution, abundance and species richness in R and BUGS. Volume 2, Dynamic and advanced models*. https://research.ebsco.com/linkprocessor/plink?id=acdec215-f169-3c05-a996-ad3f64d6341e

Kleiven, E. F., Barraquand, F., Gimenez, O., Henden, J.-A., Ims, R. A., Soininen, E. M., & Yoccoz, N. G. (2020). *A dynamic occupancy model for interacting species with two spatial scales*. https://doi.org/10.1101/2020.12.16.423067

Lahoz-Monfort, J. J., Guillera-Arroita, G., & Wintle, B. A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, *23*(4), 504–515. https://doi.org/10.1111/geb.12138

Lenoir, J., & Svenning, J.-C. (2015). Climate-related range shifts a global multidimensional synthesis and new research directions. *Ecography*, *38*(1), 15–28. https://doi.org/10.1111/ecog.00967

Lesmeister, D. B., Nielsen, C. K., Schauber, E. M., & Hellgren, E. C. (2015). Spatial and temporal structure of a mesocarnivore guild in midwestern north America: Midwestern Carnivore Guild Structure. *Wildlife Monographs*, *191*(1), 1–61. https://doi.org/10.1002/wmon.1015

Lonergan, M. (2014). Data availability constrains model complexity, generality, and utility: A response to evans *et al. Trends in Ecology & Evolution*, *29*(6), 301–302. https://doi.org/10.1016/j.tree.2014.03.005

MacKenzie, D. I., & Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*(3), 300–318. https://doi.org/10.1198/108571104X3361

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., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating Site Occupancy Rates When Detection Probabilities Are Less Than One. *Ecology*, *83*(8), 2248–2255. https://doi.org/10.1890/0012-9658(2002)083[2248:ESORWD]2.0.CO;2

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

MacKenzie, D. I., & Reardon, J. T. (2013). *Occupancy Methods for Conservation Management* (pp. 248–264). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118490747.ch11

Marescot, L., Lyet, A., Singh, R., Carter, N., & Gimenez, O. (2020). Inferring wildlife poaching in southeast Asia with multispecies dynamic occupancy models. *Ecography*, *43*(2), 239–250. https://doi.org/10.1111/ecog.04536

Mazerolle, M. J. (2017). *AICcmodavg: Model selection and multimodel inference based on (Q)AIC(c)*. https://cran.r-project.org/package=AICcmodavg

McClintock, B. T., Bailey, L. L., Pollock, K. H., & Simons, T. R. (2010). Unmodelred observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, *91*(8), 2446–2454. https://www.jstor.org/stable/27860809

McGowan, C. P., Angeli, N., Beisler, W., Snyder, C. W., Rankin, N. M., Woodrow, J., Wilson, J., Rivenbark, E., Schwarzer, A., Hand, C., Anthony, R. M., Griffin, R., Barrett, K., Haverland, A., Roach, N., Schneider, T., Smith, A. J., Smith, F., Tolliver, J., & Watts, B. D. (2020). Linking monitoring and data analysis to predictions and decisions for the range-wide eastern black rail status assessment. *Endangered Species Research*, *43*, 209–222. https://doi.org/10.3354/esr01063

Merow, C., Smith, M. J., Edwards Jr, T. C., Guisan, A., McMahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E., & Elith, J. (2014). What do we gain from simplicity versus complexity in species distribution models? *Ecography*, *37*(12), 1267–1281. https://doi.org/10.1111/ecog.00845

Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x

Miller, D., Bailey, L. L., Grant, E. H. C., McClintock, B. T., Weir, L. A., & Simons, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, *6*(5), 557–565. https://doi.org/10.1111/2041-210X.12342

Miller, D., Nichols, J., Mcclintock, B., Grant, E., Bailey, L., & Weir, L. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, *92*, 1422–1428. https://doi.org/10.2307/23035095

Miller, K. E., & Brown, J. L. (2023). Maximizing nest box monitoring effort to detect american kestrel site occupancy. *Journal of Raptor Research*, *57*(2), 176–184. https://doi.org/10.3356/JRR-22-46

Mölle, J. P., Kleiven, E. F., Ims, R. A., & Soininen, E. M. (2022). Using subnivean camera traps to study arctic small mammal community dynamics during winter. *Arctic Science*, *8*(1), 183–199. https://doi.org/10.1139/as-2021-0006

Moreira, L. F. B., Moura, R. G., & Maltchik, L. (2016). Stop and ask for directions: factors affecting anuran detection and occupancy in Pampa farmland ponds. *Ecological Research*, *31*(1), 65–74. https://doi.org/10.1007/s11284-015-1316-9

Mores, G. B., Schuler-Faccini, L., Hasenack, H., Fetzer, L. O., Souza, G. D., & Ferraz, G. (2020). Site Occupancy by Aedes aegypti in a Subtropical City is Most Sensitive to Control during Autumn and Winter Months. *The American Journal of Tropical Medicine and Hygiene*, *103*(1), 445–454. https://doi.org/10.4269/ajtmh.19-0366

Morin, D. J., Yackulic, C. B., Diffendorfer, J. E., Lesmeister, D. B., Nielsen, C. K., Reid, J., & Schauber, E. M. (2020). Is your ad hoc model selection strategy affecting your multimodel inference? *Ecosphere*, *11*(1), e02997. https://doi.org/10.1002/ecs2.2997

Nichols, J. D., Hines, J. E., Mackenzie, D. I., Seamans, M. E., & Gutiérrez, R. J. (2007). Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, *88*(6), 1395–1400. https://www.jstor.org/stable/27651247

Olson, G. S., Anthony, R. G., Forsman, E. D., Ackers, S. H., Loschl, P. J., Reid, J. A., Dugger, K. M., Glenn, E. M., & Ripple, W. J. (2005). Modeling of Site Occupancy Dynamics for Northern Spotted Owls, with Emphasis on the Effects of Barred Owls. *The Journal of Wildlife Management*, *69*(3), 918–932. https://doi.org/10.2193/0022-541X(2005)069[0918:MOSODF]2.0.CO;2

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

Otto, C. R. V., & Roloff, G. J. (2012). Songbird response to green-tree retention prescriptions in clearcut forests. *Forest Ecology and Management*, *284*, 241–250. https://doi.org/10.1016/j.foreco.2012.07.016

Padilla-Torres, S. D., Ferraz, G., Luz, S. L. B., Zamora-Perea, E., & Abad-Franch, F. (2013). Modeling Dengue Vector Dynamics under Imperfect Detection: Three Years of Site-Occupancy by Aedes aegypti and Aedes albopictus in Urban Amazonia. *PLoS ONE*, *8*(3), e58420. https://doi.org/10.1371/journal.pone.0058420

Peach, M. A., Cohen, J. B., Frair, J. L., Zuckerberg, B., Sullivan, P., Porter, W. F., & Lang, C. (2019). Value of protected areas to avian persistence across 20 years of climate and land-use change. *Conservation Biology*, *33*(2), 423–433. https://doi.org/10.1111/cobi.13205

Pendleton, D. E., Tingley, M. W., Ganley, L. C., Friedland, K. D., Mayo, C., Brown, M. W., McKenna, B. E., Jordaan, A., & Staudinger, M. D. (2022). Decadal-scale phenology and seasonal climate drivers of migratory baleen whales in a rapidly warming marine ecosystem. *Global Change Biology*, *28*(16), 4989–5005. https://doi.org/10.1111/gcb.16225

Pitman, R. T., Fattebert, J., Williams, S. T., Williams, K. S., Hill, R. A., Hunter, L. T. B., Robinson, H., Power, J., Swanepoel, L., Slotow, R., & Balme, G. A. (2017). Cats, connectivity and conservation: incorporating data sets and integrating scales for wildlife management. *Journal of Applied Ecology*, *54*(6), 1687–1698. https://doi.org/10.1111/1365-2664.12851

Pollentier, C. D., Hardy, M. A., Lutz, R. S., Hull, S. D., & Zuckerberg, B. (2021). Gobbling across landscapes: Eastern wild turkey distribution and occupancyhabitat associations. *Ecology and Evolution*, *11*(24), 18248–18270. https://doi.org/10.1002/ece3.8419

Riddell, E. A., Iknayan, K. J., Hargrove, L., Tremor, S., Patton, J. L., Ramirez, R., Wolf, B. O., & Beissinger, S. R. (2021). Exposure to climate change drives stability or collapse of desert mammal and bird communities. *Science*, *371*(6529), 633–636. https://doi.org/10.1126/science.abd4605

Risk, B. B., Valpine, P. de, & Beissinger, S. R. (2011). A robust-design formulation of the incidence function model of metapopulation dynamics applied to two species of rails. *Ecology*, *92*(2), 462–474. https://doi.org/10.1890/09-2402.1

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

Royle, J. A., & Link, W. A. (2006). GENERALIZED SITE OCCUPANCY MODELS ALLOWING FOR FALSE POSITIVE AND FALSE NEGATIVE ERRORS. *Ecology*, *87*(4), 835–841. https://doi.org/10.1890/0012-9658(2006)87[835:GSOMAF]2.0.CO;2

Scott, P. A., & Rissler, L. J. (2015). Integrating dynamic occupancy modeling and genetics to infer the status of the imperiled flattened musk turtle. *Biological Conservation*, *192*, 294–303. https://doi.org/10.1016/j.biocon.2015.10.004

Sovern, S. G., Forsman, E. D., Olson, G. S., Biswell, B. L., Taylor, M., & Anthony, R. G. (2014). Barred owls and landscape attributes influence territory occupancy of northern spotted owls. *The Journal of Wildlife Management*, *78*(8), 1436–1443. https://doi.org/10.1002/jwmg.793

Steenweg, R., Hebblewhite, M., Whittington, J., Lukacs, P., & McKelvey, K. (2018). Sampling scales define occupancy and underlying occupancyabundance relationships in animals. *Ecology*, *99*(1), 172–183. https://doi.org/10.1002/ecy.2054

Stewart, P. S., Stephens, P. A., Hill, R. A., Whittingham, M. J., & Dawson, W. (2023). Model selection in occupancy models: Inference versus prediction. *Ecology*, *104*(3), e3942. https://doi.org/10.1002/ecy.3942

Urban, M. C., Nadeau, C. P., & Giery, S. T. (2023). Using mechanistic insights to predict the climate-induced expansion of a key aquatic predator. *Ecological Monographs*, *93*(3), e1575. https://doi.org/10.1002/ecm.1575

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2023). Flexible species distribution modelling methods perform well on spatially separated testing data. *Global Ecology and Biogeography*, *32*(3), 369–383. https://doi.org/10.1111/geb.13639

Valente, J. J., Hutchinson, R. A., & Betts, M. G. (2017). Distinguishing distribution dynamics from temporary emigration using dynamic occupancy models. *Methods in Ecology and Evolution*, *8*(12), 1707–1716. https://doi.org/10.1111/2041-210X.12840

Warrier, R., Noon, B. R., & Bailey, L. (2020). Agricultural lands offer seasonal habitats to tigers in a human-dominated and fragmented landscape in India. *Ecosphere*, *11*(7). https://doi.org/10.1002/ecs2.3080

Wood, C. M., Gutiérrez, R. J., Keane, J. J., & Peery, M. Z. (2020). Early detection of rapid Barred Owl population growth within the range of the California Spotted Owl advises the Precautionary Principle. *The Condor*, *122*(1), duz058. https://doi.org/10.1093/condor/duz058

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

Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., … Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, *43*(9), 1261–1277. https://doi.org/10.1111/ecog.04960