

RESEARCH METHODS GUIDE

Estimating species occupancy across multiple sampling seasons with autologistic occupancy models via the `autoOcc` R package

Mason Fidino 

Conservation and Science Department,
Lincoln Park Zoo, Chicago, Illinois, USA

Correspondence

Mason Fidino

Email: mfidino@lpzoo.org

Funding information

Walder Foundation

Handling Editor: Laura R. Prugh

Abstract

1. Dynamic occupancy models, which estimate local colonization and extinction rates from detection/non-detection data collected across multiple sampling periods (e.g. years), are powerful but data hungry statistical tools. However, many ecological studies lack sufficient sample sizes to estimate these dynamic parameters. Autologistic occupancy models, which estimate occupancy patterns through time and account for temporal autocorrelation in a species occupancy status, offer a parsimonious alternative that is well suited for datasets with fewer sites or seasons of data.
2. Here, I introduce the `autoOcc` R package, which can be used to fit autologistic occupancy models in a frequentist framework. This package also supports model comparison via the Akaike information criterion (AIC) and making predictions from fitted models, making it a flexible and accessible option for those with detection/non-detection data collected over time.
3. Through simulations I show that autologistic occupancy models estimate parameters with less bias and more precision than dynamic occupancy models across a wide range of scenarios and sample sizes. These results suggest that autologistic occupancy models are a useful alternative when data are limited—a common constraint in ecological studies.
4. To illustrate practical use of `autoOcc` I provide two worked examples: estimating habitat associations of Virginia opossum (*Didelphis virginiana*) throughout Chicago, Illinois, USA and quantifying spatiotemporal patterns in black-backed woodpecker (*Picoides arcticus*) distributions as a function of fire severity throughout California's montane forests. These examples demonstrate not only how to implement fitting autologistic occupancy models, but also how meaningful ecological inference can be drawn from them.
5. By formally introducing this modelling framework and lowering the barrier for others to use, `autoOcc` increases the range of tools available for researchers that study species occupancy dynamics, especially when data are limited.

KEYWORDS

`autoOcc`, detection/non-detection data, detection probability, occupancy models, species distribution models

1 | INTRODUCTION

Dynamic occupancy models—which estimate a species distribution through time as a function of local colonization and extinction rates—are powerful statistical tools. These models have been used to quantify wolf (*Canis lupus*) recolonization throughout France (Louvrier et al., 2018), predict avian range dynamics (Kalle et al., 2018; Briscoe et al., 2021) and estimate pulses in colonization rates based on a species reproductive phenology (Fidino & Magle, 2017). Dynamic occupancy models are unfortunately also data-hungry (Briscoe et al., 2021; Mckann et al., 2013), which can limit how useful they are to study rare species or apply them in regions where data are limited. Consequently, there is a need for an accessible tool that effectively quantifies species distributions through time, especially with reduced sample sizes.

One approach to quantify species distributions through time is to fit a single-season occupancy model with a ‘stacked’ design (Schimpf et al., 2020). This model treats each site and primary sampling period (i.e. discrete units of time where the occupancy status of a species is assumed to be constant at each site) combination as a separate detection history and accounts for pseudoreplication via a site-level random intercept. A study with 50 sites and four primary sampling periods would have 200 detection histories if no data were missing and 50 random effect terms to estimate. Stacked design occupancy models are now relatively simple to fit because random intercepts can be incorporated into occupancy models via the unmarked R package (Fiske & Chandler, 2011). This model does not account for temporal dependence in occupancy between seasons and instead assumes that site-level occurrence per season is random and uncorrelated. Any temporal dependence in the data—which is likely present—would therefore be absorbed by some combination of the random effect term, the residual variance of the model or regression coefficients associated with covariates that vary through time, if included. As there is no way to know where the temporal dependence is absorbed, using a ‘stacked’ design could bias covariate effect estimates and hinder model interpretation. Furthermore, if there is insufficient data to fit a dynamic occupancy model, stacked design occupancy models may have the same problem, if not worse. As a ‘stacked’ design occupancy model ignores temporal dependence and requires many parameters to be estimated from detection/non-detection data, they are not likely the ideal solution to address sample size issues with data collected over time.

Autologistic occupancy models are another approach to account for spatial or temporal dependence in species distributions. Developed for spatial statistics by Besag (1974), autologistic models entered the ecological literature under that name over 20 years later (Augustin et al. 1996). In their classical form, autologistic models were used to quantify whether the occupancy probability at a site depended on the occupancy status of neighbouring sites. Wikle et al. (1998) later extended the model to include spatiotemporal dependence so that site occupancy could be influenced by the occupancy status of neighbouring sites in the previous timestep. Spatiotemporal

autologistic models have been used in a variety of ecological contexts (e.g. Hooten & Wikle, 2007) and have also been extended to dynamic occupancy models (Bled et al., 2011a, 2011b; Yackulic et al., 2012; Kase et al., 2025). More recently, however, the term autologistic has been applied to describe models in which the occupancy status of a site varies if it was occupied in the previous timestep (e.g. Tingley et al., 2016; Fidino et al., 2024). This temporal formulation is therefore equivalent to a first-order Markov process and is not explicitly autologistic because it lacks a spatial component. Nevertheless, I will retain the term ‘autologistic’ here for the temporal form to maintain consistency in the literature.

Ecologists have shown the utility of temporal autologistic occupancy models (hereafter autologistic occupancy models) across taxa, yet their overall use remains limited. This model, for example, has been used to quantify the relationship between urban mammals and gentrification (Fidino et al., 2024), between bird diversity and variation in fire severity (Tingley et al., 2016), and between amphibian occupancy and climate change (Zipkin et al., 2012). Autologistic occupancy models are a useful option because they only require one additional parameter to quantify temporal dependence in occupancy. This makes autologistic occupancy models a statistically parsimonious method for those more interested in quantifying occupancy patterns rather than the processes that inform them (Yackulic et al., 2015). Yet, despite their utility, autologistic occupancy models are seldom used because they often require researchers to write custom code to fit them, which severely limits the pool of researchers able to use this statistical method.

In this paper, I introduce the `autoOcc` R package (<https://github.com/mfidino/autoOcc>), which can be used to fit autologistic occupancy models under a frequentist framework, compare the relative fit of models via AIC and make model predictions. After explaining the basic sampling scheme and statistical formulation, I use simulations to show that autologistic occupancy models are accurate and precise under most scenarios. Furthermore, I show how `autoOcc` can be used with two worked examples: estimating habitat associations of Virginia opossum (*Didelphis virginiana*, hereafter opossum) throughout Chicago, Illinois, USA and quantifying spatiotemporal patterns in black-backed woodpecker (*Picoides arcticus*) distributions as a function of fire severity throughout California's montane forests (Stillman et al., 2023). I hope that by formally introducing autologistic occupancy models, their assumptions and making them easier to fit in R, this class of model will become a viable option for more researchers collecting species detection/non-detection data through time.

2 | EXPLANATION OF THE METHOD

2.1 | Basic sampling scheme and model assumptions

The sampling protocol for autologistic occupancy models is identical to the multi-season protocol developed by MacKenzie et al. (2003) for dynamic occupancy models. Briefly, sites are surveyed

over multiple primary sampling periods to collect species detection/non-detection data. We assume that while the occupancy status of a species may change at sites between our primary sampling periods, the occupancy status does not change within a primary sampling period (i.e. the closure assumption). During each primary sampling period, a researcher conducts multiple independent surveys at each site to generate a detection history, which is a collection of 1's and 0's that respectively indicate whether a species was or was not detected.

Autologistic occupancy models have six key assumptions in addition to the closure assumption described above. First, sampling can result in false negatives (i.e. a species is present but not detected) but not false positives (i.e. the species is not present but was mistakenly detected). Second, all sampled sites are spatially independent. Third, autologistic occupancy models assume a first-order Markov process such that the occupancy status of a site at time t conditionally depends on the preceding state at $t-1$. Temporal dependence over longer time spans (e.g. $t-3$ to t) is therefore indirectly estimated through the first-order Markov process, which lessens with increasing time span. Fourth, the probability of occupancy and detection is either constant across sites and surveys or explained by covariates. In other words, there is no unmodeled site-specific heterogeneity. Fifth, to reduce the number of parameters within the model, we assume that covariates affect colonization and persistence (i.e. 1—extinction) in identical ways. This assumption is the primary way to reduce the standard MacKenzie et al. (2003) dynamic occupancy model to the autologistic formulation (see Supporting Information S1). And sixth, if such assumptions are violated then the resulting model may be over precise or estimators could be biased and, as a result, the inference made from the associated model could be wrong (Bailey et al. 2014).

2.2 | The model

The simplest way to describe autologistic occupancy models is with a latent binary variable that denotes whether the species of interest is present or not at a site. While `autoOcc` does not use this parameterization, this is the best way to understand how the autologistic term, θ , is used in the model. Thus, for i in 1, ..., I sites and t in 1, ..., T primary sampling periods (hereafter seasons), let $z_{i,t}$ be the latent binary occupancy status of a species at site i and time t and let $\psi_{i,t}$ be the occupancy probability. During the first season there is no information about the occupancy status of the species before sampling began. Thus, when $t=1$ the latent state model is

$$\text{logit}(\psi_{i,1}) = \beta x_{i,1}$$

$$z_{i,1} \sim \text{Bernoulli}(\psi_{i,1}), \quad (1)$$

where $\beta x_{i,1}$ is the product of a vector of regression coefficients (β , which includes the model intercept) and their associated covariates ($x_{i,1}$), which is indexed by i and t because the covariates could vary across space, time or both. The first element of $x_{i,1}$ is a 1 to account for the model intercept.

For $t > 1$ we modify the logit-linear predictor by adding our autologistic term, θ , that is multiplied by a species' occupancy status at the site of interest in the previous season.

$$\text{logit}(\psi_{i,t}) = \beta x_{i,t} + \theta z_{i,t-1}$$

$$z_{i,t} \mid z_{i,t-1} \sim \text{Bernoulli}(\psi_{i,t}), \text{ for } t > 1 \quad (2)$$

When the species is present during the previous season then $z_{i,t-1} = 1$ and θ is added to the logit-linear predictor (i.e. $\beta x_{i,t} + \theta \times 1$), otherwise it is excluded (i.e. $\beta x_{i,t} + \theta \times 0$). Positive θ values indicate that species presence in the previous timestep increases $\psi_{i,t}$ in the current timestep, whereas negative θ values indicate the opposite. However, if the species was absent at $t-1$, then $\theta \times 0 = 0$, which means that θ has no bearing on the occupancy status at time t when this occurs. As such, when the autologistic θ term is added to the linear predictor it just increases or decreases the latent state model intercept, depending on its direction, given that the species is present at $t-1$.

For the detection model, let $p_{i,t,j}$ be the conditional probability of detecting the species during j in 1, ..., J secondary sampling periods (hereafter surveys) given the species is present. Further, let $y_{i,t,j}$ represent the detection/non-detection data for site i , season t and survey j which equals 1 if the species was detected, 0 if not and NA if data was not collected. This level of the model is described as,

$$\text{logit}(p_{i,t,j}) = \mathbf{a} \mathbf{k}_{i,t,j}$$

$$y_{i,t,j} \mid z_{i,t} \sim \text{Bernoulli}(p_{i,t,j} z_{i,t}) \quad (3)$$

where $\mathbf{a} \mathbf{k}_{i,t,j}$ is the product of a vector of regression coefficients (\mathbf{a}) and their associated covariates ($\mathbf{k}_{i,t,j}$) that can vary across sites, seasons or surveys. Given Equations (1) through (3), autologistic occupancy models only add one new parameter to the model, θ , to account for temporal dependence in species occupancy from one timestep to the next and add nothing to the detection-level of the model.

The latent variable approach described above could be written as a Bayesian hierarchical model in either NIMBLE (de Valpine et al., 2017) or JAGS (Plummer, 2003). However, if we drop the latent variable and write the model using marginalization with matrix notation it is possible to estimate the associated parameters via maximum likelihood, which is what I did for `autoOcc`, though this formulation could also be written in the aforementioned Bayesian software. To follow along with this model formulation, assume one site has been sampled across four seasons (indicated in square brackets) with three surveys per season to generate the detection history [110], [000], [101], [100]. Thus, the species was detected at least once in the first, third and fourth seasons and was not detected at all during the second season. For this formulation let the parameter η_t represent the occupancy probability at survey t during either the first season or when the species was not present at time $t-1$, while the parameter ω_t is the occupancy probability if the species was present at $t-1$. These two probabilities are nearly identical and can share parameters. For example, an intercept-only model for these probabilities could be written as $\text{logit}(\eta_t) = \beta_0$ and $\text{logit}(\omega_t) = \beta_0 + \theta$. Following MacKenzie et al. (2003),

and dropping the site subscript for simplicity, the probability of observing the first season detection history is

$$\Pr(\mathbf{y}_{1,1:3} = 110) = \eta_1 p_{1,1} p_{1,2} (1 - p_{1,3}). \quad (4)$$

A verbal description of Equation (4) is 'the species was present and detected on the first two surveys but not the last survey'. In the second season the species was not detected, which means either the species was not present at $t=2$ or the species was present but not detected. Because the species was present in the previous season we use ω_2 instead of η_2 . Thus, the probability of observing the second season detection history is

$$\Pr(\mathbf{y}_{2,1:3} = 000) = \left((1 - \omega_2) + \omega_2 \prod_{j=1}^3 (1 - p_{2,j}) \right). \quad (5)$$

For $t=3$, we are uncertain if the species was present in the previous season and so we use η_3 and ω_3 to generate the probability of this survey, multiplying both by the necessary detection probabilities:

$$\Pr(\mathbf{y}_{3,1:3} = 101) = (\eta_3 p_{3,1} (1 - p_{3,2}) p_{3,3} + \omega_3 p_{3,1} (1 - p_{3,2}) p_{3,3}). \quad (6)$$

A verbal description of Equation (6) is 'the species either was or was not present in season 2 and was detected on the first and third survey of season 3, but not the second survey'. Looking at Equation (6), it could be further simplified by factoring out the detection probabilities. Finally, for $t=4$ we know that the species was present in the previous timestep, so we use ω_4 :

$$\Pr(\mathbf{y}_{4,1:3} = 100) = \omega_4 p_{4,1} (1 - p_{4,2}) (1 - p_{4,3}). \quad (7)$$

The probability of observing the entire detection history across time is the product of Equations (4–7).

While it would be possible to write out the likelihood of each detection history as I did throughout Equations (4–7), doing so would be difficult to generalize to any dataset. Fortunately, MacKenzie et al. (2003) describe a more general approach that uses matrix notation for their dynamic occupancy model, which I modified for `autoOcc`. Let ϕ_0 be a row vector for the first sampling period:

$$\phi_0 = [\eta_1 \ 1 - \eta_1]. \quad (8)$$

For the remaining seasons we use a 2×2 matrix of transition probabilities, ϕ_t , that describes how a site may move from one state to the next from season t to $t+1$.

$$\phi_t = \begin{bmatrix} \omega_t & 1 - \omega_t \\ \eta_t & 1 - \eta_t \end{bmatrix}. \quad (9)$$

Note that rows of ϕ_t denote the occupancy state at t , columns are the occupancy state at $t+1$, and all rows sum to 1. For example, ω_t is the probability the species was present at t and $t+1$, while $1 - \eta_t$ is the probability the species was not present at t and $t+1$. The elements that make up ϕ_t in Equation (9) is the primary difference between autologistic and dynamic occupancy models. Finally, let $\delta_{y,t}$

be a column vector that contains the probability you would observe detection history $\mathbf{y}_{t,1:t}$ on season t , conditional on the occupancy state. While this is not how `autoOcc` handles this component of the model, it may help to imagine $\delta_{y,t}$ as a matrix of column vectors with a number of columns equal to the total number of possible detection histories. When considered in this way, $\mathbf{y}_{t,1:t}$ indexes the appropriate column vector that matches its detection history to index the correct detection probabilities. Two examples of these column vectors are

$$\begin{aligned} \delta_{110,t} &= \begin{bmatrix} p_{t,1} p_{t,2} (1 - p_{t,3}) \\ 0 \end{bmatrix} \\ \text{and} \\ \delta_{000,t} &= \begin{bmatrix} \prod_{j=1}^3 (1 - p_{t,j}) \\ 1 \end{bmatrix}. \end{aligned} \quad (10)$$

The second element of each column vector takes either the value 0 if the species was detected at least once across surveys during a season or 1 if the species was not detected. With those three components the probability of observing a given detection history is

$$\Pr(\mathbf{y}_i) = \phi_0 \prod_{t=1}^{T-1} D(\delta_{y,t}) \phi_t \delta_{y,T}. \quad (11)$$

In Equation (11), the $D(\delta_{y,t})$ function indicates that the elements in the column vector of $\delta_{y,t}$ are placed along the main diagonal of a diagonal matrix. This transformation is done to ensure that all the elements of Equation (11) are conformable and that the appropriate likelihood is calculated. Because Equation (11) is the probability of one detection history, the model likelihood is

$$\mathcal{L}(\boldsymbol{\psi}, \boldsymbol{\omega}, \mathbf{p} \mid \mathbf{y}_1, \dots, \mathbf{y}_I) = \prod_{i=1}^I \Pr(\mathbf{y}_i). \quad (12)$$

As with the dynamic occupancy model outlined in MacKenzie et al. (2003), the autologistic occupancy model can accommodate covariates via the logit link and handle missing surveys by including NA values in the detection history where data were not collected. These extensions are implemented in `autoOcc`.

2.3 | Deriving expected equilibrium occupancy estimates from autologistic occupancy models

Expected equilibrium occupancy estimates from autologistic occupancy models can be generated in a similar fashion to expected equilibrium occupancy estimates from dynamic occupancy models. As dynamic occupancy models estimate local colonization (γ) and extinction rates (ϵ), the expected equilibrium occupancy of these probabilities is $\gamma/(\gamma + \epsilon)$. By replacing those probabilities with those estimated from the autologistic occupancy model, we arrive at the following formula:

$$\eta_{i,t} / (\eta_{i,t} + (1 - \omega_{i,t})). \quad (13)$$

To demonstrate how this may look with parameters estimated via `autoOcc` and their associated covariates we could rewrite Equation (13) as

$$\text{ilogit}(\beta \mathbf{x}_{i,t}) / (\text{ilogit}(\beta \mathbf{x}_{i,t}) + (1 - \text{ilogit}(\beta \mathbf{x}_{i,t} + \theta))), \quad (14)$$

where `ilogit()` represents the inverse logit link function, $\text{ilogit}(x) = e^x / e^{1+x}$.

The estimate in Equation (13) represents the unconditional steady-state approximation of equilibrium occupancy. Conditional estimates of occupancy (e.g. occupancy given presence or absence in the previous season) can also be calculated and represent the probability of occupancy in the next season given the known occupancy status in the current season. While conditional estimates can provide more tailored estimates if a site's occupancy status is known, they require information on the prior occupancy status, which is often not available over an entire study area. If conditional estimates are of interest, however, the conditional probability of occupancy given a species absence in the previous timestep is $\text{ilogit}(\beta \mathbf{x}_{i,t})$, while if they are present, the probability is $\text{ilogit}(\beta \mathbf{x}_{i,t} + \theta)$. The `autoOcc` R package can provide any three of these three estimates via its `predict` function so that a user can select the estimate most appropriate for the goals of their study.

3 | THINGS TO CONSIDER

To provide some guidance on how much data should be collected before using `autoOcc`—and illustrate how it performs compared to other statistical techniques a researcher may use—I conducted a simulation study to compare the accuracy and precision of autologistic occupancy models to dynamic occupancy models across a range of sample sizes. For each class of model (autologistic and dynamic), simulations varied in how common or rare the species was, the number of sites sampled and the number of seasons sampled. Specifically, I used four scenarios where the expected occupancy of a species was either 0.2, 0.3, 0.4 or 0.5. For the autologistic occupancy model this was done by setting $\vartheta=1$ and using Equation (13) to determine what the model intercept should be to achieve the desired expected occupancy. For the dynamic occupancy model there are multiple colonization and extinction rates that could generate these expected occupancy probabilities. After determining possible solutions for each expected occupancy scenario I chose one at random. The number of sites sampled ranged from 30 to 100 locations in intervals of 10 while seasons sampled ranged from 4 to 12 in intervals of 2. Thus, for the four expected occupancy, seven site and six season scenarios there was a total of 160 different combinations to simulate for both the autologistic and dynamic occupancy models. Every combination was simulated and fitted 550 times and resulted in a total of 176,000 simulations and model fits.

For each simulation scenario I considered one continuous covariate on each fundamental parameter (i.e. occupancy and detection for autologistic models and initial occupancy, colonization, extinction and detection for dynamic models). Across both model

classes I set slope terms to 1 on the logit scale for all latent states except for the initial occupancy slope term of the dynamic model, which was set to 0 (i.e. the covariate was not associated to initial occupancy, though I still tried to estimate the association with the model). The initial occupancy intercept in the dynamic model was set to 1 (i.e. an initial occupancy probability of ~0.73). To model detection, the same covariate was used, but the slope term associated with this covariate was set to 0.50 across both models. Finally, I set the intercept to -0.90 on the logit scale, which resulted in a 0.29 average detection probability per survey. As I assumed four surveys per sampling period the overall probability of detecting the species at least once if they were present was about 0.75 (i.e. $1 - (1 - 0.29)^4$). In total, the autologistic and dynamic occupancy model respectively had five and eight parameters to estimate from their simulated datasets. After I fitted the simulated datasets across all scenario combinations using functions within `autoOcc`, I calculated the relative bias of each model parameter as the root mean square error (RMSE) and precision (i.e. width of 95% confidence intervals [CIs]) of latent state parameters. Across the n simulations for each scenario, the RMSE of each model parameter can be derived as $\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\tau - \hat{\tau})^2}$, where τ is the true parameter value and $\hat{\tau}$ is the estimated parameter value.

Autologistic occupancy models had less parametric bias than dynamic occupancy models. For example, all latent state parameters had an RMSE ≤ 0.5 for 53.13% of scenarios for autologistic occupancy models and 0.00% for dynamic occupancy models, with the smallest sample size to achieve this RMSE cutoff being either 70 sites with four seasons of data or 30 sites with 10 seasons. Averaged across scenarios, the relative bias of the autologistic latent state intercept was about 6.5 times less than the dynamic colonization intercept but was about 1.45 times greater than the dynamic extinction intercept (Figure 1). The autologistic term, θ , had consistently less bias which was respectively 51.0 and 5.4 times smaller than the dynamic colonization and extinction intercepts. Across all parameters, the dynamic model colonization intercept had the greatest bias when the number of sites ≤ 50 and the number of seasons ≤ 8 (Figure 1k, figure 1o). Averaged across scenarios, the relative bias in the autologistic model slope term was 4.2 times smaller than the dynamic model colonization slope term and 1.6 times smaller than the extinction slope term (Figure 2). Furthermore, the relative bias in the colonization slope term increased with the expected occupancy of the species and, like the colonization intercept, was highest when the number of sites and seasons were at their lowest across scenarios (Figure 2k).

Autologistic occupancy models were also more precise than dynamic occupancy models. Averaged across scenarios the average 95% CI width of the autologistic intercept was 1.86 times narrower than the dynamic model colonization intercept and 1.3 times narrower than the extinction intercept (Figure 3). However the autologistic term, θ , had greater uncertainty under some scenarios. The average 95% CI width for θ was roughly 1.09 times larger than the dynamic colonization intercept and 1.6 times larger than the

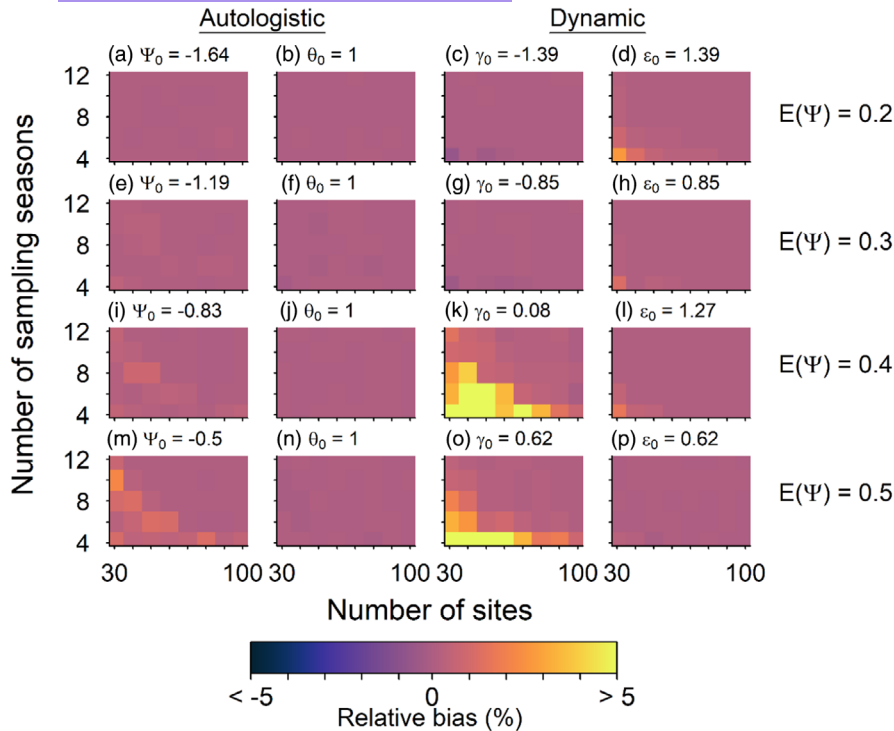


FIGURE 1 The relative bias of the latent state intercept (Ψ_0 ; a,e,i,m) and autologistic term (θ_0 ; b,f,i,n) of autologistic occupancy models as well as the colonization (γ_0 ; c,g,k,o) and extinction (ε_0 ; d,h,l,p) intercepts of dynamic occupancy models across 160 different simulation scenarios. Rows represent different scenarios where the expected occupancy of the simulated species ranges from 0.2 to 0.5, whereas the number of sites and seasons simulated varies within each subplot.

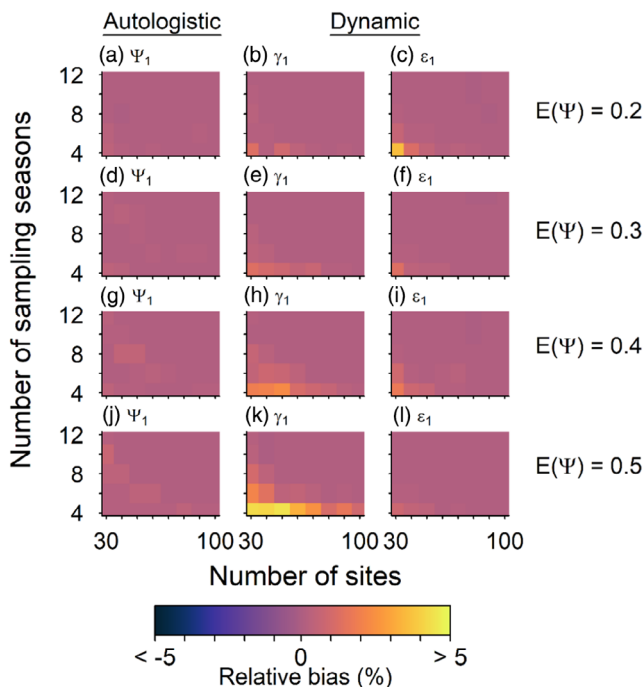


FIGURE 2 The relative bias of the latent state slope (Ψ_1 ; a,d,g,j) of autologistic occupancy models as well as the colonization (γ_1 ; b,e,h,k) and extinction (ε_1 ; c,f,i,l) slope terms of dynamic occupancy models across 160 different simulation scenarios. Rows represent different scenarios where the expected occupancy of the simulated species ranges from 0.2 to 0.5, whereas the number of sites and seasons simulated varies within each subplot.

dynamic extinction intercept. This difference was largely driven by the 0.4 and 0.5 expected occupancy scenarios, which had relatively wide 95% CI widths for both θ (Figure 3n) and the colonization intercept (Figure 3o). Averaged across scenarios, the autologistic model slope term CI was 2.67 and 1.56 times narrower than the dynamic model colonization and extinction slope term CIs, respectively (Figure 4). The largest difference between models was when the expected occupancy of the species was 0.5, especially at small sample sizes. This last result should not be surprising given that binomially distributed variables have the greatest variance when the probability of success is 0.5.

When considered together, autologistic occupancy models had less bias and more precision than dynamic occupancy models across a wider range of sample sizes. Notably, across all scenarios autologistic occupancy models had relatively low bias for all latent state parameters. This was not true for dynamic models, which especially struggled when the species was more common and when a smaller number of sites were sampled. As other researchers have suggested that dynamic occupancy models require at least 120 sites to be used reliably (McKann et al., 2013), autologistic occupancy models are clearly a valid option when dealing with small sample sizes, especially when the number of sites ≥ 70 and the number of seasons ≥ 4 . Increasing sample size did deliver a notable increase in precision for both classes of models, especially with respect to θ in the autologistic model. Most importantly, autologistic model slope terms were far more precise than dynamic model slope terms, which showcases that `autoOcc` can be especially useful if

the goal of a study is to evaluate the habitat associations of a species and a researcher is limited with the amount of data they may be able to collect.

I want to caution that the results of this simulation study cannot provide rigorous suggestions for how much data are needed before using autologistic occupancy models. The natural world is far more complex than the simulations I ran. The appropriate sample size will vary depending on your research questions, logistical constraints and the ecology of the species you plan to study. To provide some recommendations to start with, I suggest focusing on increasing precision because the relative bias was low across all scenarios of the autologistic model simulations. As such, people interested in using this class of model may be able to achieve high precision with a minimum of 70 sites sampled for eight seasons and moderate precision with 40 sites sampled for six seasons.

4 | WORKED EXAMPLES

To demonstrate how models can be fitted within *autoOcc* I have two worked examples coming from different taxa and data collection methods. For the first, I analysed camera trap data collected throughout Chicago, Illinois USA to quantify if different social-ecological gradients are associated with the distribution of the Virginia opossum (*D. virginiana*). For the second, I recreated an analysis by Stillman et al. (2023) who used a Bayesian autologistic occupancy model with survey data to assess how pyrodiversity, or the spatial and temporal

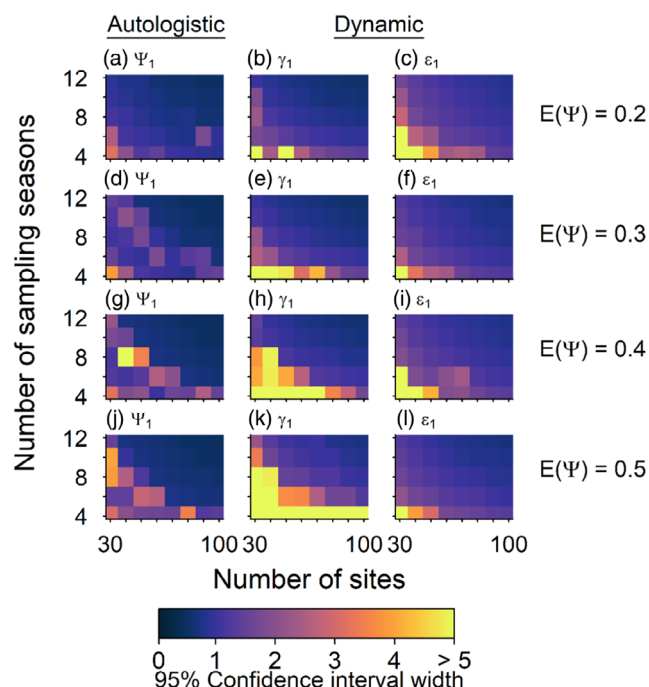


FIGURE 4 The precision (95% confidence interval width) of the latent state slope (Ψ_1 ; a,d,g,j) of autologistic occupancy models as well as the colonization (γ_1 ; b,e,h,k) and extinction (ε_1 ; c,f,i,l) slope terms of dynamic occupancy models across 160 different simulation scenarios. Rows represent different scenarios where the expected occupancy of the simulated species ranges from 0.2 to 0.5, whereas the number of sites and seasons simulated varies within each subplot.

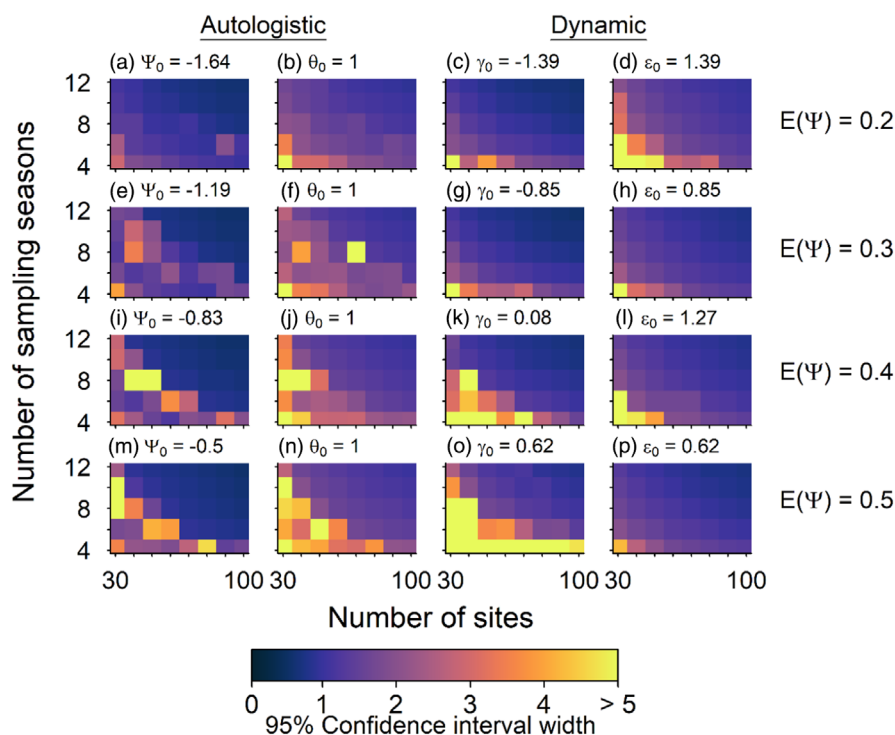


FIGURE 3 The precision (95% confidence interval width) of the latent state intercept (Ψ_0 ; a,e,i,m) and autologistic term (θ_0 ; b,f,j,n) of autologistic occupancy models as well as the colonization (γ_0 ; c,g,k,o) and extinction (ε_0 ; d,h,l,p) intercepts of dynamic occupancy models across 160 different simulation scenarios. Rows represent different scenarios where the expected occupancy of the simulated species ranges from 0.2 to 0.5, whereas the number of sites and seasons simulated varies within each subplot.

variation of fire characteristics, affects black-backed woodpecker occupancy throughout the montane forests of California, USA. Across both worked examples I compare the relative fit of different models using AIC (Anderson & Burnham, 2004), and use a Δ AIC of 2 as a cutoff value to determine which models within a model set were competitive.

4.1 | Virginia opossum occupancy throughout Chicago, IL

The data for this example comes from 96 spatial locations across the greater Chicago metropolitan area (Chicago, IL, USA). In 2019, camera traps were deployed throughout urban greenspace for 28-day sampling seasons in January, April, July and October for a total of four primary sampling periods (see Magle et al., 2019 for further sampling details). Weekly detection histories were generated from these data where one occasion represented a week of sampling. This study was approved by the Lincoln Park Zoo Research Committee (2009-028). Our internal Institutional Animal Care and Use Committee deemed this exempt from full committee review because of the non-invasive nature of this study. Permits to conduct this research were granted by the Forest Preserve District of Cook County (permit number not applicable), the Forest Preserve District of Will County (SUP-25-005), the DuPage County Forest Preserve District (21-00), Lake County Forest Preserve District (2021-133), the Chicago Park District (202322) and the Illinois Nature Preserve Commission (permit number not applicable).

For this example, we are interested in understanding how patterns of urban intensity and neighbourhood wealth are associated with opossum occupancy, both of which may be associated with opossum occupancy in non-linear ways. Furthermore, as opossum activity changes throughout the year due to Chicago's cold winters (Gallo et al., 2022), we also want to quantify the relationship between opossum detection probability and average weekly temperature.

For our analysis, we need to assemble three components in R: (1) the opossum detection history, (2) occupancy covariates and (3) detection covariates. The detection data for this example is already included within `autoOcc` so those can be loaded and set up for further analysis.

```
# load opossum detection / non-detection data
library(autoOcc)
data("opossum_det_hist")
# convert long format data to an array
opossum_y <- format_y(
  x = opossum_det_hist,
  site_column = "Site",
  time_column = "Season",
  history_columns = "Week"
)
```

This dataset has six columns that provide information on the name of the sampled location (Site), the season sampled (Season) and the detection data (Week_1 through Week_4). Furthermore, this dataset is in long format and is sorted along two columns, season and site, such that sites are sorted alphabetically within each of the four seasons of data. To use this dataset in `autoOcc`, our detection history needs to be set up as a site by season by survey three-dimensional array, which is similar to other occupancy modelling packages like `spOccupancy` (Doser et al., 2022). The `format_y()` function carries this out so long as you specify which columns denote sites, seasons and detection data. For the history columns, `format_y()` uses regular expressions to locate all columns that include whatever is included as an argument.

After setting up the detection data we need to prepare our occupancy covariates for analysis. To represent a gradient of urban intensity I used the proportion of impervious cover within 1km of each sampling location which was queried from the National Land Cover Database (Dewitz & U.S. Geological Survey, 2022). To represent neighbourhood wealth I used the median per capita income within 1km of each site from the 2014 to 2018 American Community Survey (U.S. Census Bureau, 2014–2018, 2012). These two covariates can be queried from the `opossum_covariates` dataset located within `autoOcc`. After subsetting the data I scaled the covariates for our analysis by subtracting their mean and dividing by their standard deviation, which can help improve model convergence and interpretation.

```
# load covariates
data("opossum_covariates")
# subset only impervious and income
opossum_covariates <- opossum_covariates[,
  grep("Impervious|Income", colnames(opossum_covariates))
]
# make new data.frame, scale covariates, and convert back to numeric
occ_cov_list <- opossum_covariates
occ_cov_list$Impervious <- as.numeric(scale(occ_cov_list$Impervious))
occ_cov_list$Income
<- as.numeric(scale(occ_cov_list$Income))
```

As a reminder, we wanted to quantify the relationship between opossum detection probability and average weekly temperature, which was summarized from daily temperatures provided by NCDC (National Climatic Data Center [NCDC], 2023). As temperature varies across each week of sampling, these data need to be stored in a matrix with a number of rows equal to the number of sites and a number of columns equal to the number of surveys conducted over the entire study. From there, this matrix can be stored within a named list. In our example we have 96 sites, 4 seasons of data and 4 weeks of sampling within each season so the matrix for our weekly temperature covariate will have 96 rows and 16 columns

such that the first four columns are associated to the 4 weeks of sampling in the first season and the last four columns are associated to the 4 weeks of sampling in the last season. As the temperature data I queried did not vary across space I replicated the same value along each column vector. Thus, assuming we have summarized our weather data down to 16 values, one for each week of sampling, the matrix for this detection covariate is.

```
Temperature <- matrix(
  rep(
    as.numeric(
      scale(
        weather_data_vector
      )
    ),
    each = dim(opossum_y)[1]
  ),
  nrow = dim(opossum_y)[1],
  ncol = prod(
    dim(opossum_y)[2:3]
  )
)
```

Where `weather_data_vector` is a numeric vector that contains the 16 temperature values that were queried. If we had a temporal covariate or spatiotemporal covariate that only varied by primary sampling period and not surveys we could instead store those data in a site by season matrix. Our temperature matrix, as well as any other covariates we may want to include on opossum detection probability, are stored within a named list. For example, if I wanted to control for our two social-ecological gradients on detection probability they can be included like so.

```
det_cov_list <- list(
  Temperature = Temperature,
  Impervious = occ_cov_list$Impervious,
  Income = occ_cov_list$Income
)
```

With these three pieces of data you can use `autoOcc` to fit a suite of models with the `auto_occ()` function and then compare their relative fit with `compare_models()`. For this analysis I fitted 10 models. Models understandably varied in which covariates were included but also whether those covariates had a quadratic term. For example, there are eight possible models that could be fitted with two occupancy covariates, both with and without quadratic terms. The last two models included a temperature only model (i.e. intercept only for occupancy, but temperature and temperature² on detection probability) and a null model. Every model except for the null model included temperature and temperature² in the detection logit-linear predictor. As an example, the global model could be specified as

```
global_quadratic <- auto_occ( ~Temperature +
  I(Temperature^2) +
  Impervious + I(Impervious^2) + Income +
  I(Income^2)
  ~Impervious + I(Impervious^2) + Income +
  I(Income^2),
  y = opossum_y,
  det_covs = det_cov_list,
  occ_covs = occ_cov_list
)
```

Note that the model formulas are the first argument of this function, and just as with the `unmarked` package, are written as a double right-hand side formula for detection and occupancy, in that order. After fitting the remaining models and storing them in a list you can compare their relative fit with the `compare_models()` function via AIC. There was only one competitive model which included a linear effect of income on opossum occupancy but a quadratic effect of impervious cover, and we can use the `summary()` function to look at the estimated parameters of this model.

```
# What the model list could look like after fitting the 10
models
```

```
model_list <- list(
  global_quadratic = global_quadratic,
  global = global,
  income_quadratic = income_quadratic,
  income_quad_imperv = income_quad_imperv,
  income = income,
  imperv_quadratic = imperv_quadratic,
  imperv_quad_income = imperv_quad_income,
  imperv = imperv,
  temperature = temp,
  null = null
)
```

```
# compare models via AIC
```

```
aic_results <- compare_models(
  model_list,
  digits=2
)
```

```
# Look at first few models
```

```
head(aic_results, 3)
  model npar  AIC delta AICwt cumltvWt
imperv_quad_income  11 1213.11  0.00  0.55  0.55
  global_quadratic  13 1215.27  2.16  0.19  0.74
  imperv_quadratic   9 1216.16  3.05  0.12  0.86
```

```
# Summarise best-fit model
```

```
summary(imperv_quad_income)
```

```
Call:
```

```
autoOcc::auto_occ(formula = ~Temperature +
  I(Temperature^2) +
```

```

Impervious + I(Impervious^2) + Income ~ Impervious +
I(Impervious^2) +
Income, y = opossum_y, det_covs = det_cov_list, occ_
covs = occ_cov_list)
optim convergence code: 0
optim iterations: 51
Occupancy estimates:
parameter Est SE lower upper p
psi - (Intercept) -0.324 0.200 -0.716 0.0674 1.05e-01
psi - Impervious -0.341 0.165 -0.664 -0.0177 3.87e-02
psi - I(Impervious^2) -0.274 0.131 -0.531 -0.0180
3.60e-02
psi - Income -0.279 0.164 -0.601 0.0436 9.02e-02
psi - theta 1.637 0.365 0.922 2.3517 7.16e-06
Note: psi - theta is the autologistic term
Detection estimates:
parameter Est SE lower upper p
rho - (Intercept) 0.3826 0.145 0.0984 0.6667 0.008315
rho - Temperature 0.3374 0.133 0.0761 0.5986 0.011373
rho - I(Temperature^2) -0.4384 0.116 -0.6661 -0.2106
0.000162
rho - Impervious -0.0536 0.145 -0.3386 0.2315 0.712669
rho - I(Impervious^2) -0.1291 0.110 -0.3446 0.0863
0.240025
rho - Income -0.2054 0.130 -0.4601 0.0493 0.113935
AIC: 1213.113

```

Making predictions for this model is relatively straightforward via the `predict()` function in `autoOcc`. Assume we want to predict opossum occupancy across a gradient of impervious cover, which ranged from about 20% to 80% across this study. These predictions can be made in three steps. First, we generate a dataset with covariate values we'd like to make predictions with, holding other covariates at their mean. As we mean-centred our continuous covariates we can keep all other columns at their mean value of 0.

```

# A vector of impervious cover
imperv_vec <- seq(20, 80, length.out=300)
# The prediction data.frame
imperv_dm <- data.frame(
  matrix(
    0,
    ncol = ncol(imperv_quad_income@occcovs),
    nrow = length(imperv_vec)
  )
)
# add column names
colnames(imperv_dm) <- names(imperv_quad_income@
occcovs)

```

Second, our impervious cover data needs to be scaled exactly as we for our analysis. Thus, we center and scale the prediction

impervious cover vector by the mean and standard deviation of the impervious cover data at our sampled sites.

```

# add in imperv and scale it in the same way as we
did in the model
imperv_dm$Impervious <- (
  imperv_vec - mean(opossum_covariates$Impervi-
ous)
) / sd(opossum_covariates$Impervious)

```

Finally, we can make predictions. At a minimum the `predict()` function in `autoOcc` requires you to provide the model you'd like to make predictions with and the type of prediction you'd like to make ('psi' for the latent state, 'rho' for detection). You can also add a new dataset to make predictions with and input the confidence level you would like for CIs. Finally, because this function uses Monte Carlo simulations to approximate CIs you can also specify a sufficiently large number of Monte Carlo simulations (defaults to 3000) and a seed.

```

Imperv_pred <- predict(
  object = imperv_quad_income,
  type = "psi",
  newdata = imperv_dm,
  nsim=3000,
  seed=453
)

```

These predictions can then be plotted out (Figure 5). In this example, opossum occupancy was greatest when impervious cover was roughly 43% and decreased with increasing levels of neighbourhood wealth. I also detected a non-linear relationship between opossum detection probability and weekly temperature such that opossum detection probability was highest when the weekly average temperature was around 15°C (Figure 5).

4.2 | Black-backed woodpecker occupancy throughout the montane forests of California

For this worked example I recreated the analysis conducted by Stillman et al. (2023), who used a Bayesian autologistic occupancy model to quantify how pyrodiversity affects black-backed woodpecker occupancy. For this study these authors had an impressive 10 years of woodpecker survey data at nearly 2400 sites, though sites were not surveyed each year. For their sampling protocol, they used single-visit surveys that were divided into multiple independent surveys. Specifically, each survey consisted of playing black-backed woodpecker vocalizations for 30s and then listening for responses for 1.5 min. A maximum of three playback surveys were conducted at each site, and playback surveys were stopped after the first detection (i.e. a removal design). Finally, point counts were

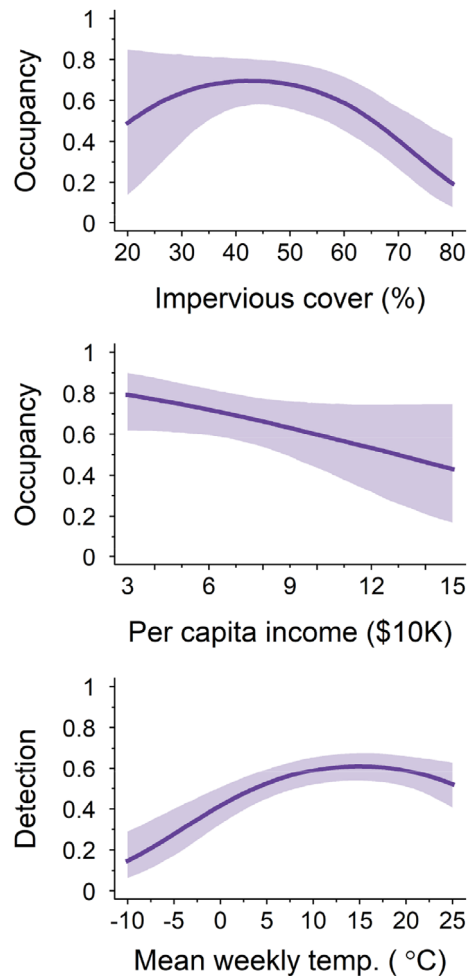


FIGURE 5 Opossum occupancy throughout Chicago, Illinois varied non-linearly with impervious cover and linearly with per capita income whereas opossum detection probability had a non-linear relationship with mean weekly temperature. The detection/non-detection data used to generate this figure are supplied with `autoOcc` and came from a long-term camera trapping study throughout Chicago.

also conducted at roughly half of these sites after the playback surveys. For further information on sampling see Tingley et al. (2018, 2020).

Stillman et al. (2023) compared three models: a static occupancy model, a temporal autologistic occupancy model and a landscape-temporal autologistic occupancy model. These last two models varied in which covariates were present within the model. While Stillman et al. (2023) also incorporated random effects into their models, `autoOcc` cannot. As such, random effects were not used for this worked example.

I fitted three models to these data. All models had the same detection logit-linear predictor, which included whether a survey was 2 or 3 min in length (2 min = 0, 3 min = 1), the ordinal day of a survey and the survey type (point count = 0, playback = 1). As a departure from Stillman et al. (2023), I did not fit a static occupancy model as the first model and instead used an intercept-only autologistic

model. The second model, which Stillman et al. (2023) described as their 'temporal occupancy model', accounted for the effects of years since fire on occupancy. This model included five occupancy covariates: the number of years since a fire had started at a site, the average burn severity of the last fire within 100m, the proportion pre-fire canopy cover within 100m, elevation and latitude. This model also included quadratic terms for elevation and fire age as well as a statistical interaction between elevation and latitude. The third, 'landscape-temporal occupancy model', had the covariates from the temporal occupancy model but also included other covariates to account for post-fire habitat dynamics. As such, in addition to the aforementioned covariates this model included a metric for diversity in burn severity, distance to a low burn severity forest patch and the summed basal area of pre-fire live red fir and white fir within 100m. This landscape-temporal occupancy model included quadratic terms for elevation and fire age as well as statistical interactions between burn severity and fire age, elevation and fire age, and fir basal area and fire age.

Just like Stillman et al. (2023), I found that the landscape-temporal occupancy model had the best relative fit with the temporal occupancy model having a ΔAIC of 122.48 from the landscape-temporal model. Parameter estimates from the best-fit model were similar to the Bayesian analysis of Stillman et al. (2023) such that inference made across models is comparable (Figure 6). However, because I did not include random effects in the autologistic occupancy model there was also less parametric uncertainty (i.e. smaller CIs) with the `autoOcc` analysis across every model parameter. Regardless, had this analysis originally been conducted with `autoOcc`, Stillman et al. (2023) would have found similar results and concluded that woodpecker occupancy was highest directly after a fire, especially if the site had a high burn severity (Figure 6). For the complete analysis in R, see Supporting Information S1.

5 | CAVEATS

One of the main strengths of autologistic occupancy models is that temporal dependence in site-level occupancy is accounted for with a single parameter. This makes the model simple and easier to apply to datasets that are smaller. However, this simple formulation may also be a weakness as it inherently assumes that θ does not vary across seasons or sites. Certainly, spatiotemporal covariates can be included within the model, but that does not explicitly quantify whether site-level persistence (i.e. the probability a site remains occupied if a species is present there in the previous timestep) varies across space or time. Autologistic occupancy models could be extended to address this specific issue, but at that point such a model is nearly equivalent to a dynamic occupancy model. Therefore, as with any analysis, it is important to consider what your research questions or management goals are to ensure that the model used is appropriate.

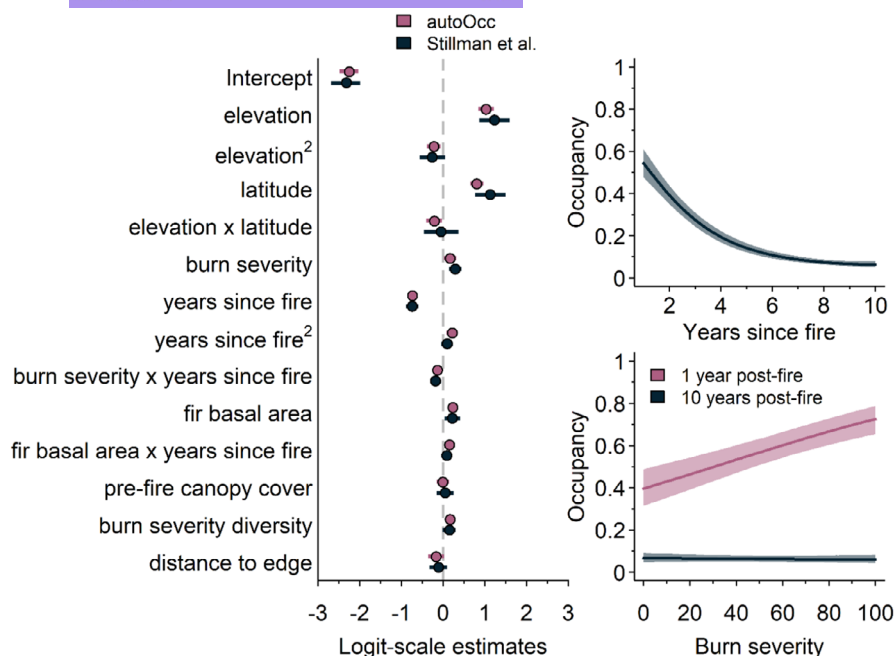


FIGURE 6 A comparison of autologistic occupancy model parameter estimates of the same dataset analysed under a frequentist framework in autoOcc or a Bayesian framework in JAGS (Stillman et al., 2023). Though the models fitted were not identical (i.e. the Bayesian analysis included random effects), similar inferences can be made from the resulting model parameters. Median parameter estimates and 95% credible intervals of the Stillman et al. (2023) analysis were collected from the supporting information of their manuscript. The right two plots represent occupancy predictions generated from autoOcc as a function of years since a fire occurred as well as the relationship between years since fire and burn severity.

6 | ADDITIONAL RESOURCES

In addition to this manuscript, `autoOcc` has an associated vignette that can be accessed so long as it is built when installing `autoOcc`. This can be viewed by running `vignette("Overview")` in R after loading the package. This vignette provides a description of autologistic occupancy models, how to add spatial, temporal or spatiotemporal covariates to your autologistic occupancy models and shows how all the functions within `autoOcc` can be used to run a complete analysis on the opossum dataset.

7 | CONCLUSION

Biodiversity monitoring surveys that collect data through time are increasingly common, yet the statistical tools to quantify habitat associations with such data have been somewhat limited when said data are subject to imperfect detection. The `autoOcc` package offers up perhaps the simplest approach that accounts for temporal dependence, and through simulations I demonstrated how autologistic occupancy models can have relatively little bias even with small sample sizes. Through worked examples I provided a general outline for how an analysis can be conducted within the `autoOcc` ecosystem and then recreated the results of an already published study assessing black-backed woodpecker responses to fire severity (Stillman et al., 2023). Ultimately, I hope that all the information and results presented here, in combination with the `autoOcc` R package, will be helpful to researchers who are interested in understanding species occupancy patterns over time.

AUTHOR CONTRIBUTIONS

Mason Fidino was the sole author of this research.

ACKNOWLEDGEMENTS

I am grateful to Andrew Stillman and Morgan Tingley for their help while I was working on recreating the woodpecker analysis. I also thank Seth Magle and Brian Gerber for their comments on a previous draft of this manuscript, as well as two anonymous reviewers. The reviewers were especially helpful in providing historical context on autologistic occupancy models and showing how the autologistic occupancy model is a special case of a dynamic occupancy model.

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code for the simulation and case studies available from the Zenodo repository <https://doi.org/10.5281/zenodo.17467340> (Fidino, 2025a). Code for the `autoOcc` R package available from the Zenodo repository <https://doi.org/10.5281/zenodo.17064781> (Fidino, 2025b).

STATEMENT ON INCLUSION

The main product of this manuscript is an R package and so this section is not really applicable. I helped collect the data for one of the worked examples, whereas the other dataset was publicly available (and I consulted with the authors who generated that dataset to ensure I modelled it in a similar fashion). Those scientists have been listed in the acknowledgements.

ORCID

Mason Fidino  <https://orcid.org/0000-0002-8583-0307>

REFERENCES

- Anderson, D., & Burnham, K. (2004). *Model selection and multi-model inference* (Vol. 63, p. 10). Springer-Verlag.
- Augustin, N. H., Muggleston, M. A., & Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33, 339–347.
- 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.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 36(2), 192–225.
- Bled, F., Royle, J. A., & Cam, E. (2011a). Hierarchical modeling of an invasive spread: The Eurasian collared-dove *Streptopelia decaocto* in the United States. *Ecological Applications*, 21(1), 290–302.
- Bled, F., Royle, J. A., & Cam, E. (2011b). Assessing hypotheses about nesting site occupancy dynamics. *Ecology*, 92(4), 938–951.
- Briscoe, N. J., Zurell, D., Elith, J., König, C., Fandos, G., Malchow, A. K., & Guillerá-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.
- de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., & Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2), 403–413.
- Dewitz, J., & U.S. Geological Survey. (2022). *National land cover database (NLCD) 2019 products, version 2.0*. U.S. Geological Survey. <https://doi.org/10.5066/P9KZCM54>
- Doser, J. W., Finley, A. O., Kéry, M., & Zipkin, E. F. (2022). spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, 13, 1670–1678.
- Fidino, M., Sander, H., Lewis, J., Lehrer, E. W., Rivera, K., Murray, M. H., Adams, H. C., Kase, A., Flores, A., Stankowich, T., Schell, C. J., Salsbury, C. M., Rohnke, A. T., Jordan, M. J., Green, A. M., Gramza, A. R., Zellmer, A. J., Williamson, J., Surasinghe, T. D., ... Magle, S. (2024). Gentrification drives patterns of alpha and beta diversity in cities. *Proceedings of the National Academy of Sciences of the United States of America*, 121(17), e2318596121.
- Fidino, M. (2025a). mfidino/autologistic_walkthrough (v1.0). Data from: Estimating species occupancy across multiple sampling seasons with autologistic occupancy models via the autoOcc R package. *Zenodo*. <https://doi.org/10.5281/zenodo.17467340>
- Fidino, M. (2025b). mfidino/autoOcc (V0.1.2). *Zenodo*. <https://doi.org/10.5281/zenodo.17064781>
- Fidino, M., & Magle, S. B. (2017). Using Fourier series to estimate periodic patterns in dynamic occupancy models. *Ecosphere*, 8(9), e01944.
- Fiske, I., & Chandler, R. (2011). Unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10), 1–23.
- Gallo, T., Fidino, M., Gerber, B., Ahlers, A. A., Angstmann, J. L., Amaya, M., Concilio, A. L., Drake, D., Gay, D., Lehrer, E. W., Murray, M. H., Ryan, T. J., St Clair, C. C., Salsbury, C. M., Sander, H. A., Stankowich, T., Williamson, J., Belaire, J. A., Simon, K., & Magle, S. B. (2022). Mammals adjust diel activity across gradients of urbanization. *eLife*, 11, e74756.
- Hooten, M. B., & Wikle, C. K. (2007). Invasions, epidemics, and binary data in a cellular world. In *Proceedings of the American Statistical Association* (pp. 3999–4010).
- Kalle, R., Ramesh, T., & Downs, C. T. (2018). When and where to move: Dynamic occupancy models explain the range dynamics of a food nomadic bird under climate and land cover change. *Global Change Biology*, 24(1), e27–e39.
- Kase, A., Fidino, M., Lehrer, E. W., & Magle, S. B. (2025). Local and long-distance colonization influence the distribution of a species in a fragmented landscape. *The Stacks*, 1–32.
- Louvrier, J., Duchamp, C., Lauret, V., Marboutin, E., Cubaynes, S., Choquet, R., Miquel, C., & Gimenez, O. (2018). Mapping and explaining wolf recolonization in France using dynamic occupancy models and opportunistic data. *Ecography*, 41(4), 647–660.
- 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.
- Magle, S. B., Fidino, M., Lehrer, E. W., Gallo, T., Mulligan, M. P., Ríos, M. J., Ahlers, A. A., Angstmann, J., Belaire, A., Dugelby, B., Gramza, A., Hartley, L., MacDougall, B., Ryan, T., Salsbury, C., Sander, H., Schell, C., Simon, K., St Onge, S., & Drake, D. (2019). Advancing urban wildlife research through a multi-city collaboration. *Frontiers in Ecology and the Environment*, 17(4), 232–239.
- Mckann, P. C., Gray, B. R., & Thogmartin, W. E. (2013). Small sample bias in dynamic occupancy models. *The Journal of Wildlife Management*, 77(1), 172–180.
- National Climatic Data Center [NCDC]. (2023). *National oceanic and atmospheric administration*. National Climatic Data Center. Retrieved June 25, 2023, from <https://www.ncdc.noaa.gov/cdo-web/datatools/finds/tation>
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 124(125.10), 1–10.
- Schrimpf, M. B., Che-Castaldo, C., & Lynch, H. J. (2020). Regional breeding bird assessment of the Antarctic peninsula. *Polar Biology*, 43(2), 111–122.
- Stillman, A. N., Wilkerson, R. L., Kaschube, D. R., Siegel, R. B., Sawyer, S. C., & Tingley, M. W. (2023). Incorporating pyrodiversity into wildlife habitat assessments for rapid post-fire management: A woodpecker case study. *Ecological Applications*, 33(4), e2853.
- Tingley, M. W., Ruiz-Gutiérrez, V., Wilkerson, R. L., Howell, C. A., & Siegel, R. B. (2016). Pyrodiversity promotes avian diversity over the decade following forest fire. *Proceedings of the Royal Society B: Biological Sciences*, 283(1840), 20161703.
- Tingley, M. W., Stillman, A. N., Wilkerson, R. L., Howell, C. A., Sawyer, S. C., & Siegel, R. B. (2018). Cross-scale occupancy dynamics of a post-fire specialist in response to variation across a fire regime. *Journal of Animal Ecology*, 87(5), 1484–1496.
- Tingley, M. W., Stillman, A. N., Wilkerson, R. L., Sawyer, S. C., & Siegel, R. B. (2020). Black-backed woodpecker occupancy in burned and beetle-killed forests: Disturbance agent matters. *Forest Ecology and Management*, 455, 117694.
- U.S. Census Bureau, 2014–2018. (2012). *American community survey 3-year public use microdata samples* (2021). U.S. Census Bureau. Retrieved June 25, 2023, from <https://www.census.gov/programs-surveys/acs/data.html>
- Wikle, C. K., Berliner, L. M., & Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5(2), 117–154.
- Yackulic, C. B., Nichols, J. D., Reid, J., & Der, R. (2015). To predict the niche, model colonization and extinction. *Ecology*, 96(1), 16–23.
- Yackulic, C. B., Reid, J., Davis, R., Hines, J. E., Nichols, J. D., & Forsman, E. (2012). Neighborhood and habitat effects on vital rates: Expansion of the Barred Owl in the Oregon Coast Ranges. *Ecology*, 93(8), 1953–1966.

Zipkin, E. F., Grant, E. H. C., & Fagan, W. F. (2012). Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecological Applications*, 22(7), 1962–1972.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supporting Information S1.

How to cite this article: Fidino, M. (2025). Estimating species occupancy across multiple sampling seasons with autologistic occupancy models via the `autoOcc` R package. *Journal of Animal Ecology*, 00, 1–14. <https://doi.org/10.1111/1365-2656.70184>