Estimating species occupancy through time with the `autoOcc` R package.

# Introduction

# Explanation of the method

The simplest way to describe this class of statistical model is with a latent binary variable that denotes whether the species of interest is present or not at a sampled location. While autoOcc does not use this parameterization to estimate the associated model parameters we explain it here to provide intuition into how the autologistic term, *θ*, is used. Thus, for *i* in 1,…,*I* sites and *t* in 1,…,*T* primary sampling periods, let *zi,t* be the occupancy status of a species at site *i* and time *t* and *ψi,t* be the occupancy probability. During the first timestep there is no information about the occupancy status of the species before sampling began. Thus, when t=1the latent state model is

|  |  |  |
| --- | --- | --- |
|  | *, t = 1* | Eq. 1 |

where is a vector of regression coefficients (including the model intercept) and their associated covariates, which are indexed by *i* and *t* because they could vary across space, time, or both. Note that the first element of is a 1 to account for the model intercept. This setup for the first season of data is identical to dynamic occupancy models which estimate initial occupancy in the first timestep. However, autologistic occupancy models use the parameters that describe a species occupancy in the first timestep across all seasons instead of separately estimating local colonization and extinction rates. 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 timestep.

|  |  |  |
| --- | --- | --- |
|  | *, t > 1* | Eq. 2 |

Thus, when the species is present during the previous timestep and *θ* is added to the logit-linear predictor, otherwise it is excluded. Positive *θ* values indicate that species presence in the previous timestep increases *ψi,t* in the current timestep, whereas negative *θ* values indicate the opposite.

For the data model let *p* be the conditional probability of detecting the species during *j* in 1,…,*J* secondary sampling periods given the species presence. Further, let yi,t,j represent the detection / non-detection data for site *i*, primary sampling period *t*, and secondary sampling period *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 then

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 3 |

where is a vector of regression coefficients and their associated covariates that may vary across sites, primary, and secondary sampling periods. Given Eq. 1 through Eq. 3, the autologistic parameterization adds one new parameter to the model, *θ*, to account for temporal dependence in species occupancy from one timestep to the next.

Instead of describing this model with an underlying latent variable, which is most commonly done in a Bayesian analysis, it is possible to construct the likelihood of observing each datapoint instead by modifying the original dynamic occupancy model outlined in MacKenzie et al. (2003). To provide some intuition for how the autologistic model can be formulated in this way, let’s assume a site has been sampled across four primary sampling periods with three repeat surveys to generate the detection history of 110, 000, 101, 100. Thus, the species was detected at least once on the first, third, and fourth primary sampling periods and was not detected at all during the second. To ensure we don’t confuse the separate occupancy probabilities, let ψ

WRITING HERE

Following MacKenzie et al. (2003), the probability of observing the first sampling period detection history is

describe the model in this way I assume that probabilities do not vary across space or time, and as such remove their associated subscripts, but as with the previous parameterization all the associated probabilities can be made a function of covariates with the logit link. Furthermore, for the sake of example, let’s assume a site has been sampled across four primary sampling periods with three repeat surveys to generate the detection history of 110, 000, 101, 100. Thus, the species was detected at least once on the first, third, and fourth primary sampling periods and was not detected at all during the second. Following MacKenzie et al. (2003), the probability of observing the first sampling period detection history is:

|  |  |  |
| --- | --- | --- |
|  |  | Eq. 4 |

Psi\_theta 1 – psi\_theta

Psi 1 - psi

## Deriving expected occupancy estimates from this model

# Things to consider before using this method

## Underlying model assumptions

## Sample size considerations

# Worked examples

Bit of text. First about data coming being provided within autoOcc. Camera trap data. Second, a renalaysis of Stillman et al. (2023), that had point data data across FIRE STUFF. Unique because they only visited sampling locations once per year, but split that single visit into multiple secondary sampling periods.

## Virginia opossum throughout Chicago, IL

## Black-backed woodpecker throughout California

# Caveats

# Additional resources

# Conclusion