Setting Thresholds: Optimal detection probabilities for reliable multi-state occupancy models

Thomas Osinga

Master thesis 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Wildlife, Fish, and Environmental Studies
Freestanding course
Examensarbete / SLU, Institutionen för vilt, fisk och miljö
2024:12
Umeå 2024

Setting Threshold: Optimal detection probabilities for reliable multi-state occupancy models

Thomas Osinga

Supervisor:	**Dr. Tim Hofmeester, SLU, Wildlife, Fish and Environmental Studies**
Assistant supervisor:	Dr. Henjo de Knegt, WUR, Wildlife Ecology and Conservation Group
Assistant supervisor:	Dr. Magali Frauendorf, SLU, Wildlife, Fish and Environmental Studies
Examiner:	Dr. Sheila Holmes, SLU, Wildlife, Fish and Environmental Studies

Credits:	30
Level:	A2E
Course title:	Master thesis in Biology, A2E
Course code:	EX0971
Programme/education:	Freestanding Course
Course coordinating dept:	Department of Wildlife, Fish, & Environmental Studies
Place of publication:	Umeå,
Year of publication:	2024
Cover picture:	Camera trap images from the 'Beyond Moose' project
Copyright:	All featured images are used with permission from the copyright owner
Title of series:	Examensarbete / SLU, Institutionen för vilt, fisk och miljö
Part number:	2024:12

Keywords: Multi-state, Hierarchical modelling, Camera trap, Bayesian, Moose, Roe Deer

Swedish University of Agricultural Sciences
Faculty of Forest Sciences
Department of Wildlife, Fish, and Environmental Studies

Abstract

Camera trapping has become crucial in wildlife research, enabling detailed observations of elusive and nocturnal species with limited human interference. The use of occupancy modelling to analyse camera trap data is also rapidly increasing, aiding in the assessment of habitat selection, species distribution, and multi-species dynamics while considering imperfect detection. However, the design of camera trap studies, typically involving large grids with limited cameras, often results in sparse data and low detection probabilities which challenges model accuracy. These low detection probabilities can hamper model performance and resulting inferences, however to what extent this is true for multi-state occupancy models is still unknown. In this study, I simulated datasets with 60 and 200 sites, with various detection probabilities between 0.001 and 0.9 and their impact on multi-state occupancy models to establish thresholds required for model reliability. The results revealed that detection probabilities must exceed 0.1 for model convergence, and that values above 0.2 minimized bias and eliminated convergence issues. Additionally, more than 60 sites are required in multi-state occupancy models for finding habitat relationships, which is more than found for single-state occupancy models. The case study on moose and roe deer revealed that while aggregation of survey length increases detection probabilities it also increases uncertainty in estimates. Models with daily detection histories failed to converge, while weekly aggregation improved fit for some parameters, however there still had large uncertainty. Simplifying state complexity enhanced model performance despite low detection probabilities but could not fully mitigate the effects of the sparse data. I conclude that when detection probabilities are too low, complex occupancy models are difficult to fit. Strategies like deploying multiple cameras, targeted camera placement or using bait can increase detection rates but may introduce biases. However, the gained model performance from higher detection probabilities might outweigh these biases. Ensuring higher detection probabilities and adequate data collection is essential for reliable model outcomes. This study highlights key thresholds and considerations for improving multi-state occupancy models using camera trap data, aiding in the design of more effective wildlife research studies.

Table of contents

1. Introduction

Camera trapping has become a common technology in wildlife research, enabling detailed observations of species in their natural habitats without the need for direct human interaction (Delisle et al., 2021; Fisher, 2023). This technology has proven particularly valuable in the study of elusive and nocturnal species, where traditional observation methods fall short or are too cost-intensive (Bondi et al., 2010; Hofmeester et al., 2024; Palmeirim et al., 2019). Among the many applications of camera trap data, occupancy modelling stands out as a flexible methodological approach to assess habitat selection and species distribution (Rovero & Zimmerman, 2016). This technique has seen a significant increase in popularity, largely in response to earlier oversights where many did not account for imperfect detection (Burton et al., 2015). However, fitting occupancy models to camera trap data introduces several challenges that must be addressed. These challenges arise from the limitations in effort per location due to the finite number of cameras, the duration of their deployment, and the restricted detection distance and view angle inherent to camera traps (Burton et al., 2015; McIntyre et al., 2020; O'Connor et al., 2017). Consequently, the resulting low detection probabilities (the likelihood of detecting a species, given it is present) can affect the accuracy of detecting and estimating the occupancy states of species (Kéry & Royle, 2020; Pautrel et al., 2024; Steenweg et al., 2019). These inherent issues illustrate the need for a deeper understanding of how low detection probabilities impact the reliability of occupancy models and the robustness of ecological inferences that can be drawn from such studies.

Occupancy models are designed to estimate the probability that a given 'site' is used or occupied by a species while considering the imperfect detection of individuals (MacKenzie et al., 2017; Mackenzie & Royle, 2005). This approach provides a robust framework for making inferences about species distribution patterns based on detection-non-detection data collected over time (Kéry & Royle, 2020). However, applying these models to data from camera trap studies, which often employ an extensive grid layout, introduces significant challenges (Rovero & Zimmerman, 2016; Royle & Dorazio, 2008). In these studies, camera traps are distributed across larger grid cells, each containing one or more sampling locations (Burton et al., 2015). Plot sizes are often arbitrarily chosen because they do not correspond well to the actual field of view of the camera, which is often much smaller than the grid cell (as discussed in Efford & Dawson, 2012 & Steenweg, 2018). The chosen plot sizes are then often much larger than the area a camera effectively samples, leading to sparse datasets with infrequent repeat detections at occupied sites across survey dates. Additionally, for species with larger home ranges, this results in high occupancy estimates because occupancy is interpreted as the 'use' of a location, and these species use multiple grid cells, albeit rarely (Steenweg et al., 2018). When occupancy of a study area is high and detection probabilities are low, it becomes challenging for the model to ascertain the true occupancy status of sites where the species was not detected, increasing the uncertainty of ecological assessments (Efford & Dawson, 2012; Steenweg et al., 2019).

Low detection probabilities have been shown to be an issue for model-fitting within occupancy models (Delisle et al., 2023; Mckann et al., 2013; Pautrel et al., 2024; Steenweg et al., 2019). Even for the simplest models, low detection probabilities can severely compromise the accuracy and reliability of model fitting. Simulation studies for simple occupancy models have shown that detection probabilities below 0.2 can result in biases and convergence issues (Kéry & Royle, 2020). To increase detection probability one can aggregate continuous camera trap data into longer survey periods (Burton et al., 2015), which reduces temporal resolution but increases detection probability and theoretically should have limited impact on occupancy estimates (Rovero & Zimmerman, 2016). However, reaching a detection probability of 0.2 can still be challenging when studying a species which is elusive or in low densities and when there are limited number of cameras per site. The current knowledge of model functioning and detection probability thresholds is all based on single state static and dynamic occupancy models but for more complex model specifications (e.g. multi-state, multi-species, and for models including covariates) this has yet to be explored.

Low detection probabilities could be particularly problematic for multi-state occupancy models (including multi-species and community models; Rota et al., 2016), which are designed to capture various occupancy states of a species at a site (e.g., unoccupied, occupied without offspring, occupied with offspring, Nichols et al., 2007). While these models are versatile and offer nuanced interpretations of ecological systems, their effectiveness hinges on the accurate classification of states. This accuracy, in turn, depends on the frequency and reliability of detection events. Multi-state models also require the estimation of extra parameters, necessitating a substantially larger dataset (Kéry & Royle, 2020). Therefore, it is important to consider these factors when planning a study. Insufficient detection probabilities or an inadequate number of sites or surveys can severely limit any model's inference capabilities. While I expect this to be a problem for multi-state occupancy models it is unknown if multi-state occupancy models have different data requirements than the frequently studied single species models.

In this thesis, I aim to identify the required detection probabilities and implications of using occupancy models with camera trap data, specifically focusing on how low detection probabilities affect the accuracy and reliability of these models in ecological assessments. I will simulate multiple scenarios for detection probabilities using a fixed number of sites and surveys and evaluate the model outcome to the true values used for simulation. Additionally, I will conduct a case study with data on moose and roe deer to illustrate convergence issues and the effects of different data aggregations, highlighting the challenges associated with camera trap data aggregation and the detection probability requirements for stable model convergence. This simulation study serves as a starting point for designing future studies that utilize camera traps for (multi-state) occupancy modelling, I highlight the considerations and adjustments needed for applying these models effectively in wildlife research. Furthermore, the case study is used as an example of the possibilities and limitations associated with very limited data.

.

2. Methods

2.1 Multi-state occupancy modelling framework

2.1.1 Model structure

The model used for simulation was structured as a hierarchical, single-season multi-(four)-state occupancy model akin to the model described by Mackenzie et al. (2006). The simulation study is based on a study system of Moose, focusing on the occupancy of female moose and their reproductive status during summer (May to September). Figure 1 gives a schematic visualization of the dependencies between the different possible occupancy states.

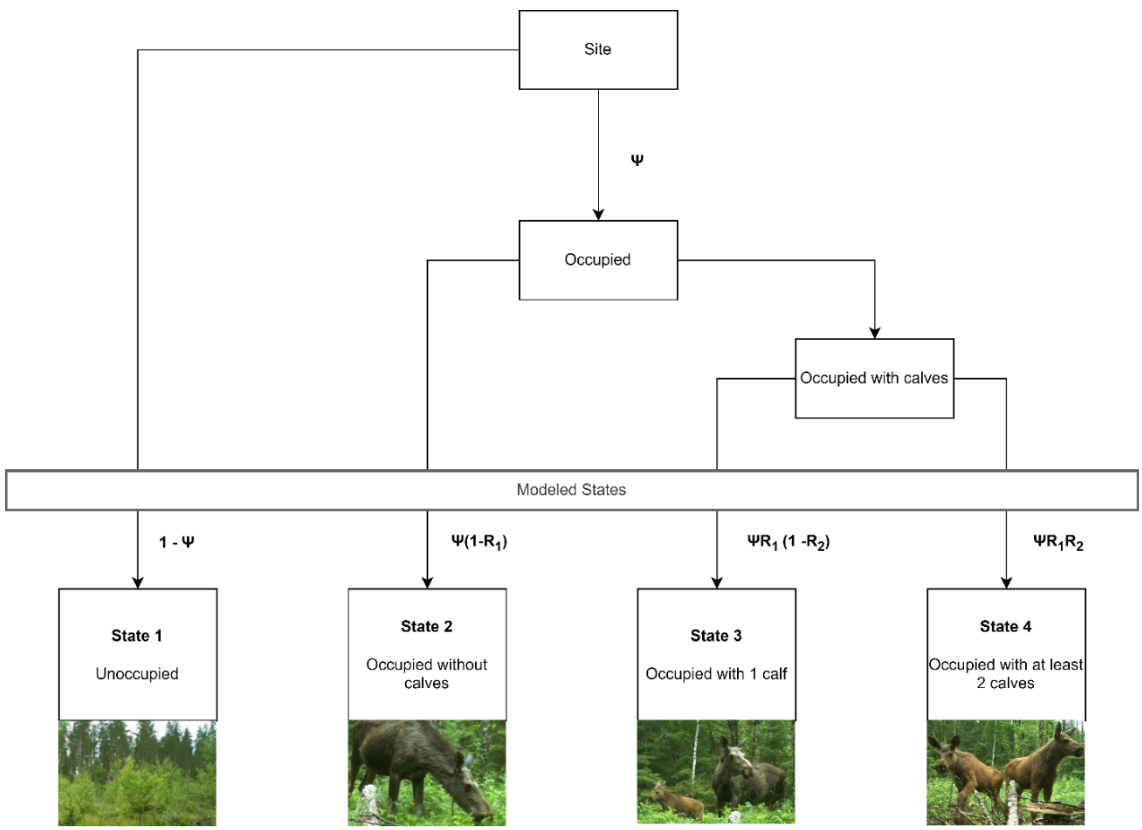

Figure 1: Diagram of the structure of the multi-state occupancy model used in this study, indicating dependencies and probability parameters. Ψ is indicating the probability of site occupancy, whereas $1 - \Psi$ indicates the probability of a site not being occupied. Occupancy is defined as location is 'used' during the study period. The formulas indicate the conditional probability of the state, given that the site is occupied (see text for elaboration).

The model considers a total of four states, with one state representing unoccupied sites and the other three representing occupied sites with varying reproductive outcomes. The probability of site occupancy (Ψ) is the foundational layer, upon which the reproductive states are conditional. The R coefficients determine the relationship between the states.

$\boldsymbol{\Omega}_s[m]$: $Probability\ of\ each\ state\ per\ site$

$\boldsymbol{\Psi} = \boldsymbol{Pr}(Occupied) = \Omega_s[2] + \Omega_s[3] + \Omega_s[4]$

$\Omega_s[1] = \boldsymbol{Pr}(Not\ occupied) = 1 - \Psi$

$\Omega_s[2] = \boldsymbol{Pr}(\,Occupied\ without\ calves \mid Occupied\,) = \Psi_s(1 - R_s^1)$

$\Omega_s[3] = \boldsymbol{Pr}(\,Occupied\ with\ 1\ calf \mid Occupied\,) = \Psi_s\, R_s^1(1 - R_s^2)$

$\Omega_s[4] = \boldsymbol{Pr}(\,Occupied\ with\ > 1\ calves \mid Occupied\,) = \Psi_s\, R_s^1\, R_s^2$

In this simulation study I modelled the occurrence of a species across a network of camera traps sampled over a period of time to understand its distribution and detection patterns. Specifically, I conducted a series of surveys across distinct sites to collect data. For each site *s*, ranging from 1 to *S* sites, I have carried out *J* surveys, where *j* spans from 1 to *J* and stored their observations in an *S* by *J* matrix ***Y.*** The latent true state Z_s, was modeled at each site during each survey using a categorical distribution, with the probabilities of each state defined by **Ω** (*eq. 2.1, and eq. 2.2*).

$$\boldsymbol{\Omega_s} = [(1 - \Psi_s),\ \Psi_s(1 - R_s^1)\,,\ \Psi_s\, R_s^1(1 - R_s^2),\ \Psi_s\, R_s^1\, R_s^2\,] \quad (eq.\ 2.1)$$

$$Z_s \sim Categorical\big(\boldsymbol{\Omega}_s^i\,\big) \quad (eq.\ 2.2)$$

The influence of environmental and site-specific covariate on state occupancy ($\boldsymbol{\Omega_s}$) was modelled using hierarchical multinomial logit functions (*eq. 2.3, 2.4, and 2.5). These relationships* are represented by the equations:

$$logit(\Psi_s) = \beta_0 + \beta_1 * X_{1,s} + \cdots + \beta_r * X_{r,s} \quad (eq.\ 2.3)$$

$$logit(R1_s) = \beta_0 + \beta_1 * X_{1,s} + \cdots + \beta_r * X_{r,s} \quad (eq.\ 2.4)$$

$$logit(R2_s) = \beta_0 + \beta_1 * X_{1,s} + \cdots + \beta_r^1 * X_{r,s} \quad (eq.\ 2.5)$$

Where $X_{r,s}$ represents the covariate matrix and *r* indicates the specific covariate for each site s. $\beta_{r,s}$ are the coefficients estimating the influence of these covariates. $\beta_{0,r}$ is the intercept for each model providing a baseline of state occupancy.

Detection probability (Θ) was kept constant across sites and surveys for simplicity. The detection matrix (*eq.* 2.6) presents the probability of detection based on the latent state *z[i],* where each row represents an observed state, and each column corresponds to a modelled latent state (both ordered from state 1 to 4). The values within the matrix reveal the probability of detecting an observed state given the modelled latent state. This matrix is built with the assumption that false positives are not possible, hence everything on the right side of the matrix is filled with zeros. This assumption is thought to be adhered to as every image is manually classified and antlers of bulls have already started forming in the beginning of May making the chance of false positives to be in the data extremely low.

$$\Theta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 - p^{1,1} & p^{1,1}0 & 0 & 0 \\ 1 - (p^{2,1} + p^{2,2}) & p^{2,1} & p^{2,2} & 0 \\ 1 - (p^{3,1} + p^{3,2} + p^{3,3}) & p^{3,1} & p^{3,2} & p^{3,3} \end{bmatrix} \qquad (eq. 2.6)$$

Here, $p_s^{z,j}$ represents the probability of detection given the latent state, with z indicating the initial latent reproductive state and j specifying the observed state.

$$\boldsymbol{y}_{s,j} \sim Categorical(\Theta_{\boldsymbol{s}}, z_{[s]}) \qquad \textit{(eq. 2.7)}$$

The latent state $z_{[s]}$ and the accompanying detection probabilities (Θ) are then combined to model the observed state through equation 2.7.

The full commented model code can be found on GitHub[1]. Here you also find multiple parametrizations such as a three-state model (A model simplification where state 4 and 3 are combined), and models with covariates for occupancy or/and detection probabilities.

2.2 Simulation and power analysis

2.2.1 Simulation design

This simulation study was designed to predict occupancy probabilities of moose across a landscape under a forest density gradient. I simulated data for 60 sites over 35 survey dates, following recommendations by Kays et al. (2020) to achieve both reliable covariate relationships and accurate occupancy estimates. The chosen site number (40-60) is considered a realistic minimum for a study period of 4-5 weeks (28-35 days).

I modelled ten different detection probabilities ranging from 0.001 to 0.9 to detect the correct occupancy state ($p^{1,1}$, $p^{2,2}$, $p^{3,3}$), while misclassification probabilities were held constant at 0.05 ($p^{2,1}$, $p^{3,1}$, $p^{3,2}$; Table 1). For each detection probability, I generated 100 datasets with no covariates, and 260 datasets with one covariate, each sampled with the 200 unique true states sampled from the occupancy matrix $\boldsymbol{\Omega}_{s,state}$ (*eq. 2.1*) and the detection matrix $\Theta_{p[i]}$ (*eq. 2.6 & eq 2.7*).

---

[1] https://github.com/ThomasOsinga/MultiStateOccupancyPowerAnalysis

The occupancy probabilities ($\mathbf{\Omega}$) for the model without covariates were determined by fixed parameters for Ψ (0.7), $R_s^1(0.5)$ and $R_s^2(0.2)$. For the simulation with covariates, the occupancy probabilities for each state were modeled using a linear model on the logit scale, applied to a uniformly distributed site covariate, forest density, which I scaled from -2 to +2 (*eq. 2.3, 2.4 & 2.5*). Fixed regression coefficients were used for the linear model (Table 1). Each simulation retained the same sequence of 'true states' across different detection probability settings to ensure consistency.

Table 1: Initial values used for each detection probability and intercept and slope parameters for models including one covariate for the different simulations.

Parameters	**Values**
Detection Probability (p11, p22, p33)"	0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7, 0.9
Misclassification Probability (p21, p31, p32)"	0.05
$\beta_{0,[1,1]}$	log(3)
$\beta_{0,[2,1]}$	log(0.5)
$\beta_{0,[3,1]}$	log(0.1)
$\beta_{1,[1,1]}$	4
$\beta_{1,[2,1]}$	5
$\beta_{1,[3,1]}$	1

2.2.2 Statistical Modelling Using JAGS

The simulated data were analysed using JAGS (Just Another Gibbs Sampler) in R utilizing the jagsUI package (v.1.6.2, Kellner 2024). JAGS is a software designed for Bayesian hierarchical modelling via Markov Chain Monte Carlo (MCMC) methods using the BUGS language (Bayesian inference Using Gibbs Sampling). I developed a Bayesian multi-state occupancy model in BUGS to estimate the relationship between forest density and the multiple states of moose occupancy, considering both direct detections and misclassification errors. I ran the model for each dataset, where the posterior distributions of the parameters were estimated over 50,000 iterations, with the first 32,500 iterations discarded as burn-in with a thinning rate of 5 to avoid autocorrelation. I ran the model in parallel across 3 chains. Convergence diagnostics were assessed using the Gelman-Rubin statistic ($\hat{R}$) and effective sample size. Model specifications, including priors for the fixed effects (i.e., intercept and slope for forest density) and the detection probabilities, are fully detailed in the model script available in Appendix V).

The posterior distributions were analysed by calculating the bias, coverage (meaning the proportion of true values falling within the 95% credible intervals of the posterior), root mean square error (RSME) and, coefficient of variation (CoV) for each detection probability scenario across the simulations. The percentage of model convergence was noted as a percentage of models having ($\hat{R}$) values below 1.1 and above 0.9 for all parameters.

2.3 Case study: moose and roe deer in northern Sweden

2.3.1 Study aim

To test the limits of the multi-state occupancy model I used a dataset from a camera trap study in northern Sweden. Population densities of mammals are relatively low in this part of the world, so I tested the model on the two species with the most sightings during the study period. Roe deer and moose were the only two species in my data meeting the requirements for our model, mostly being solitary during summer and having the ability to have multiple offspring. These two requirements ensure that the same state is unlikely to occur on the same location, adhering to the conditional hierarchical structure of the model.

2.3.2 Camera trap data wrangling

To analyse differences in habitat selection between females without offspring and those with one or multiple offspring, I used camera trap data collected from a study conducted in the Järnäshalvön peninsula in northeast Sweden, as part of the Beyond Moose project (Cromsigt et al., 2023). The details of the study design can be found in Hofmeester et al. (2020). The camera traps were in operation from the spring of 2017 until the spring of 2018 however, for this study I specifically included data from May 2017 to September 2017. This timeframe was selected because I focused on the habitat selection of females with their offspring. Before May, few offspring are born (Hofmeester et al., 2020; Solberg et al., 2007), and after August, the lactation period is concluded for most individuals (Mauget et al., 1997, 1999; Severud et al., 2019). The data is annotated to species level, including the distinction between sex (female, male, unknown) and life stage (adult, sub-adult, juvenile, unknown). I reclassified the data into three classes, occupied female (adult and sub-adult female moose), occupied with a single calf, and occupied with multiple calves.

For this study, I transformed the data to a detection-non detection data format. This transformation incorporates a temporal dimension, aggregating data in each location as detected or non-detected within a timeframe, defining the survey length. The survey length was set to daily, as this still holds most of the information, and weekly to see if increased detection probabilities give better model estimates.

2.3.3 Modelling occupancy and covariate integration

I ran four and three-state occupancy models (as detailed in 2.1 and Appendix V) with one covariate and without covariates to evaluate the limits of my data on model convergence. I ran the models over 100,000 iterations, with the first 60,000 iterations discarded as burn-in with a thinning rate of 5 to avoid autocorrelation. I ran the model in parallel across three chains. Model performance was analysed using the Gelman-Rubin statistic ($\hat{R}$), effective sample size, confidence intervals and, viewing posterior distribution and traceplots.

The model was structured to estimate four states for each site: unoccupied, occupied without offspring, occupied with one calf, and occupied with multiple calves. The detection process was modelled separately, accounting for potential misclassification of the true occupancy state during each survey. For the three-state model I simplified the model where all sightings with 1 and more than 1 calf were counted as state 3. Full commented model code for the four-state model can be found in Appendix V, the three state model is found on GitHub[2].

[2] https://github.com/ThomasOsinga/MultiStateOccupancyPowerAnalysis

3. Results

3.1 Simulation results

3.1.1 Four-state occupancy model with no covariates

Fitting the occupancy model without covariates to simulated data reveals that the model begins to converge to the true value when the detection probability exceeds 0.1 (Figure 2). This phenomenon is evident in all fitted parameters, as variability in the means decreases with increasing detection probability, with a distinct threshold at 0.1. Additionally, fewer than 5% of the models fail to converge when the detection probability exceeds 0.1 (Appendix I, Figure 1a). However, the standard deviation across all fitted parameters only drastically decreases when detection probabilities reach 0.2 or higher (Appendix I, Figure 1b).

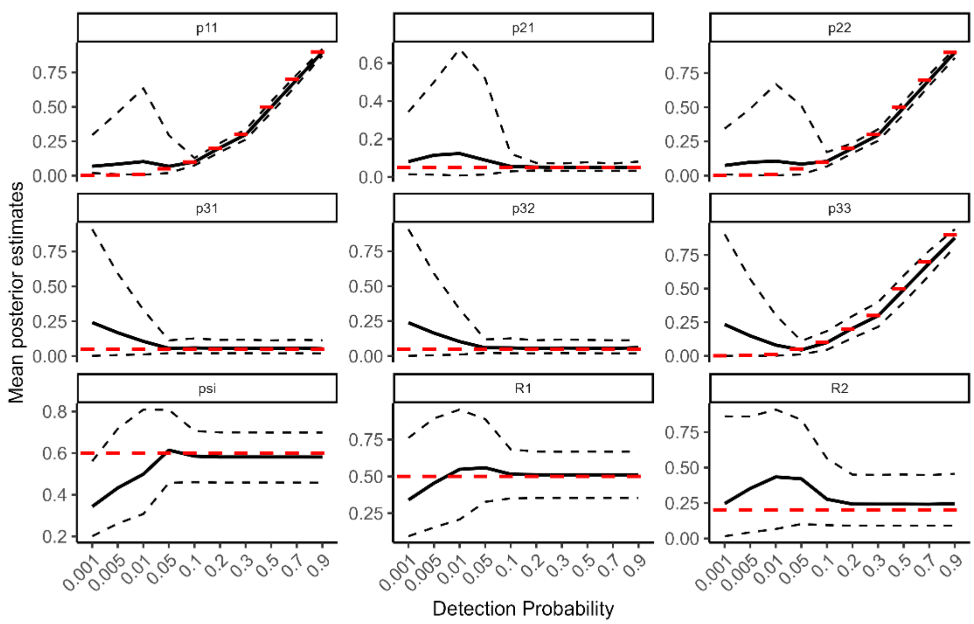

Figure 2: The average posterior estimates for the mean and 95% credible intervals across 100 simulations from the four-state model without covariates are shown for each fitted parameter (y-axis) across 10 different simulation scenarios with varying detection probabilities (x-axis). The black dashed lines represent the average upper and lower credible intervals, while the solid black lines represent the average mean posterior estimate. The horizontal red lines indicate the true parameter values used in the simulations. See the model description (chapter 2.1) for an explanation of the different parameters.

The diagnostic analysis of simulations with the basic model without covariates reveals that bias, coefficient of variability (CoV), and root mean square error (RMSE) follow a consistent pattern across all fitted parameters. Specifically, the values for bias, RMSE and, CoV are elevated when the detection probability is below 0.05 and begin to decrease and stabilize at a detection probability of 0.2 (Appendix I, Figures 2, 3 & 4). Coverage, the number of simulations where the true value falls between the credible intervals, remains above 0.75 for all models with detection probabilities above 0.005 and exceeds 0.90 for detection probabilities above 0.05 (Appendix I, Figure 5)

3.1.2 Four-state model with one covariate

Fitting the occupancy model without covariates to simulated data mirrors the model without covariates when it goes for detection and occupancy estimates (Figure 3). All models converge from a detection probability of 0.2, the means stabilize, and a decrease in standard deviation is found (Appendix II, Figure 1a & 1b). However, the regression parameters are more difficult to fit, for many simulations the relationships between the covariate and the states are not found, even for high detection probabilities. Especially the coefficient for R_1 ($\beta_{1,[2,1]}$) was often not estimated close to the true value (see Appendix II, Figure 2 for a more detailed graph). The coefficient for R_2 ($\beta_{1,[3,1]}$) is relatively close to zero and crosses zero in multiple simulations, indicating that the effect is also often not found (Appendix II, Figure 2a).

The diagnostic analysis of the four-state model with one covariate aligns closely with the model without covariates (see Appendix II, Figures 4, 5 & 6). Depending on the parameter model performance increases greatly between 0.05 and 0.2. Meaning bias, CoV, and RMSE drop significantly. However, two coefficients show a high bias, CoV and RMSE across all simulation scenarios. The slope coefficients associated with occupancy (Ψ) and occupied with at least one calf (R_1) exhibit high bias, RMSE, and CoV. Coverage remains high (>75%) for all parameters but is less informative at lower detection probabilities due to highly uncertain estimates (see Appendix II, Figure 3).

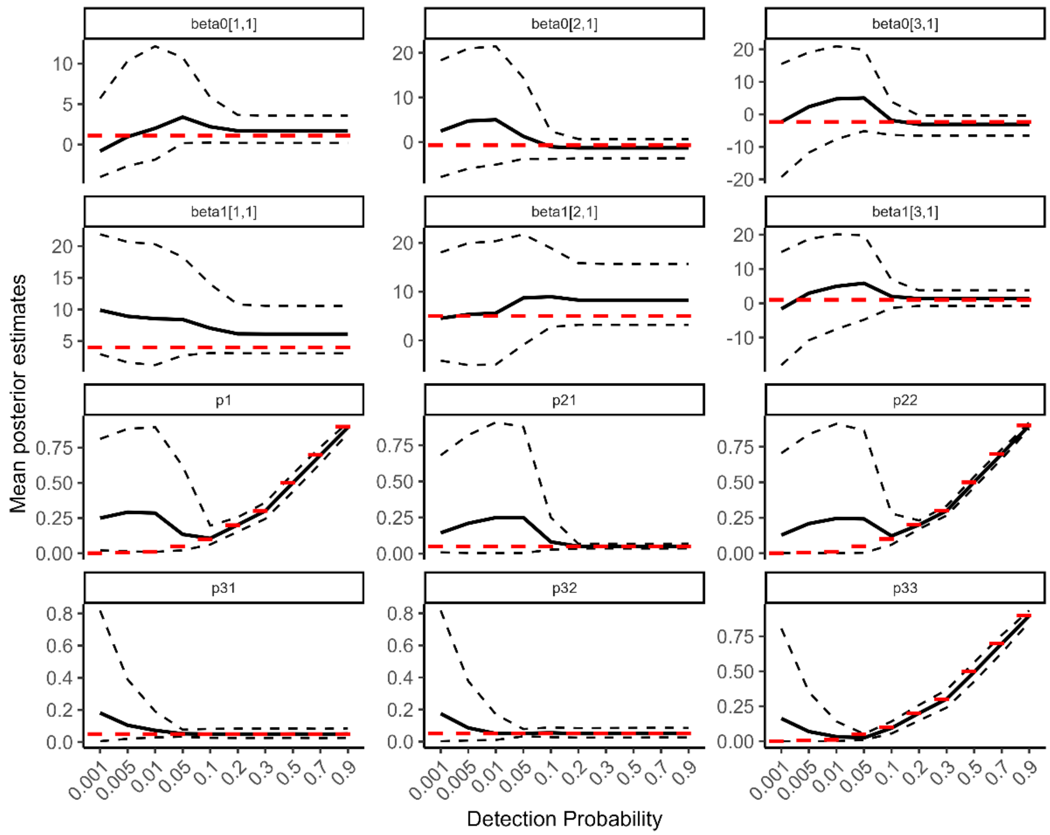

Figure 3: The average posterior estimates for the mean and 95% credible intervals across 260 simulations from the four-state model with one covariate are shown for each fitted parameter (y-axis) across 10 different simulation scenarios with varying detection probabilities (x-axis). The black dashed lines represent the average upper and lower credible intervals, while the solid black line represents the average mean posterior estimate. The horizontal red line indicates the true parameter values used in the simulations. See the model description (chapter 2.1) for an explanation of the different parameters.

The model with 200 sites instead of 60 showed clear improvements in estimating beta coefficients. Bias was reduced from 3-5 to less than 1 at detection probabilities above 0.2 (see Appendix III, Figure 1). Additionally, the thresholds for detection probability parameters decreased from between 0.1 and 0.2 to between 0.05 and 0.1, enhancing the model's performance on covariate relationships. This indicates that increasing the number of sites drastically improves the accuracy and reliability of parameter estimates.

3.2 Case Study

3.2.1 Four-state static model with no covariates

Daily detection history

Fitting a four-state occupancy model to daily survey data did not give accurate estimates. Multiple parameters when using daily survey length had $\hat{R}$ values above 1.1 and confidence intervals covering most of the parameter space. For moose the model did converge; however, model estimates were still very uncertain which is seen by the large credible intervals (Figure 4).

Weekly detection history

When fitting the models using weekly survey length for both roe deer and moose almost all parameters converged, with the exception for $p^{1,1}$ (the probability of detecting an adult with no calves given there was an adult with no calves). However, credible intervals were large and increased for all parameters compared to the models with daily survey lengths, except for $p^{1,1}$ for roe deer and R_2 for moose, for which the credible intervals decreased drastically.

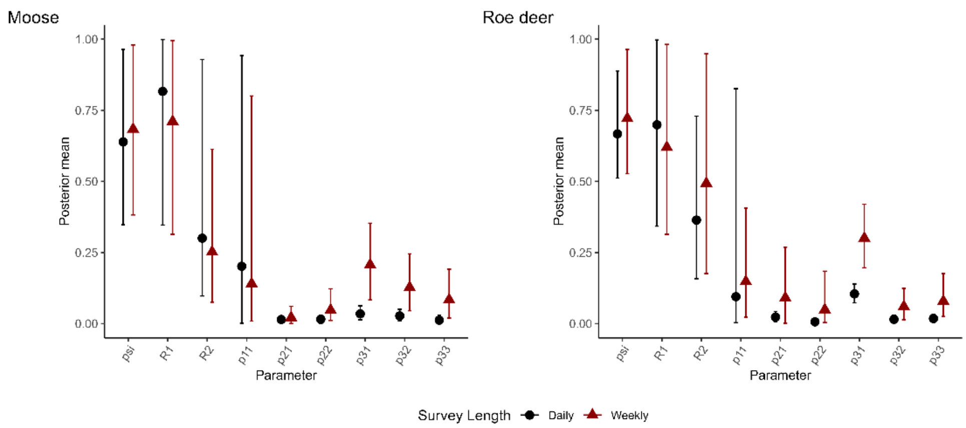

Figure 4: The posterior mean estimate and 95% credible intervals (y-axis) of both the moose and roe deer four-state occupancy model. The detection and occupancy parameters are shown on the x-axis. See the model description (chapter 2.1) for an explanation of the different parameters.

3.2.2 Three-state static model with no covariates

Daily and weekly detection histories

Fitting two three-state occupancy models with daily and weekly detection history both converged for all parameters and had acceptable confidence intervals for roe deer (Figure 5). For moose, the occupancy estimates (Ψ, r) and $p^{1,1}$ converged but estimates were extremely uncertain as seen in credible intervals ranging between 0 and 1 (Figure 5). The model with weekly detection probability as input had slightly larger confidence intervals. This was the case for both moose and roe deer models.

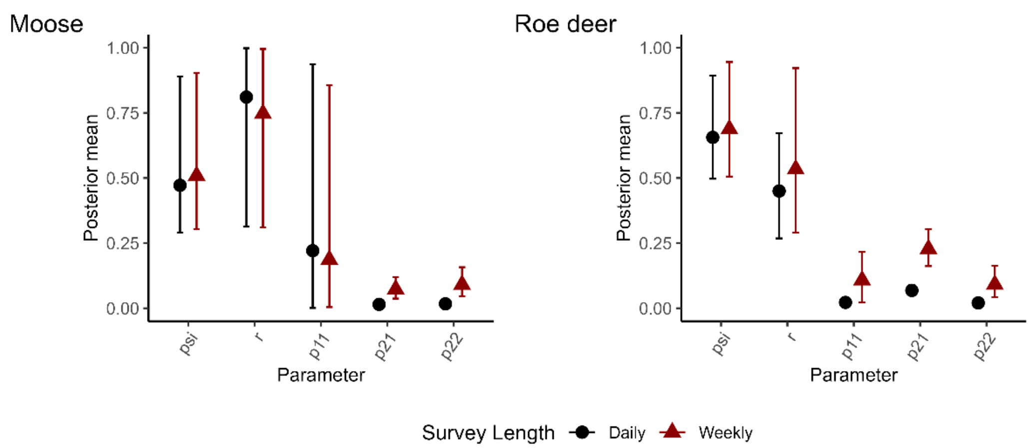

Figure 5: The posterior mean estimate and 95% credible intervals (y-axis) of both Moose and Roe deer three-state occupancy model. The detection and occupancy parameters are shown on the x-axis. See the model description (chapter 2.1) for an explanation of the different parameters.

3.3 Three and four-state static models with one covariate

None of the models incorporating a covariate converged, therefore I took no inferential conclusions from them (Appendix IV, Figures 1 & 2).

4. Discussion

In this study, I simulated various scenarios to explore a range of detection probabilities, from low to high, and their impact on fitting (multi-state) occupancy models. These simulations aimed to establish minimum detection probability thresholds necessary for model confidence. My findings suggest that detection probabilities must exceed 0.1 for high model confidence, with values above 0.2 being preferable to minimize bias, reduce variability, and alleviate convergence issues.

Although data aggregation to increase survey length and, in turn, detection probability can address model convergence issues associated with low detection probabilities, this approach is not a remedy for data insufficiency. Aggregating data by reducing the number of surveys increases the uncertainty in estimates of detection probabilities and, consequently, the occupancy parameters, ultimately affecting the robustness of model outputs. When covariates were added model performance went down drastically, indicating that more than the 60 sites used for simulation are required to make inferences when using multi-state models when there is no strong prior knowledge of the relationship. These findings highlight the importance of ensuring sufficiently high detection probabilities in multi-state occupancy models. Therefore, to achieve robust and reliable model outputs, I recommend prioritizing high detection probabilities and an adequate number of sites, especially when incorporating covariates without strong prior knowledge. Despite these challenges, my study successfully identifies key thresholds and considerations for improving model performance for multi-state occupancy models.

Camera traps provide continuous detection data within a limited field of view, which researchers often use to infer wildlife presence in a broader area. The size of the area and the method of data aggregation significantly influence both the interpretation and the fitting of models. Daily aggregation for survey length, the most common approach for camera trap data in occupancy models (Burton et al., 2015; Rovero & Zimmerman, 2016), typically results in low detection probabilities, particularly in regions with sparse species populations. Model performance was closely linked with detection probabilities, showing similar patterns as found in single-state models (Mckann et al., 2013; Pautrel et al., 2024). Pautrel et al. (2024) noted that detection probabilities need to reach at least 0.1 for adequate model fit. It's also important to recognize that the required detection probability is influenced by the occupancy probability (Mackenzie & Royle, 2005), which was held constant at around 0.7 in this study. Contrary to what you might expect, when occupancy increases a higher detection probability is required to get confident estimates (Mackenzie & Royle, 2005; Steenweg et al., 2019). This implies that in areas with lower occupancy, detection probabilities below 0.2, could still yield functional results. However, an occupancy of 0.7 is quite common for species caught by camera traps (e.g. Hofmeester et al., 2021; Rich et al., 2016; Steenweg et al., 2019; Wevers et al., 2021), thus suggesting that these results are widely applicable for most multi-state occupancy models.

In occupancy studies that are not bound by time detection probabilities might not significantly limit the analysis. Research on single-species occupancy and capture-recapture models indicate that when the cumulative probability of detecting a species at least once during the survey period (p^*) exceeds 0.8, occupancy estimates remain reliable (Smith et al., 2009; Steenweg et al., 2019). However, achieving this detection probability can be challenging for elusive species with high occupancy or low abundance and especially when your study period is limited. Moreover, while single-state occupancy models only need to account for detection probability, multi-state models must also account for the potential misclassification of states. The complexity of multi-state models means they demand a significant amount of data to be reliably fitted. This need for abundant data can pose a challenge, especially when dealing with rare or elusive species. Therefore, despite the potential advantages of multi-state models, their application necessitates careful consideration of data availability and the specific challenges associated with detecting and classifying different states in wildlife populations.

Furthermore, besides detection probability, the number of surveys and sites plays a crucial role in study design, particularly when incorporating covariates. Previous research suggested that 60 sites and 35 survey days might be sufficient to establish covariate relationships (Kays et al., 2020). However, my findings for multi-state models challenge this assertion, none of the models based on 60 sites in this study accurately estimated the slope coefficients. My simulation using 200 sites shows great improvement in slope coefficient estimates, but estimates at high detection probabilities are still not always certain (Appendix III, Figure 1a). This is consistent with findings by Kéry & Royle (2020), who were unable to fit a multi-species model even with a subsample of 140 sites from an original dataset of 1400, despite finding clear relationships with the full dataset. Additionally, successful modelling efforts involving multi-state, species, or community occupancy models incorporating covariates typically involve between 100-2000 sites (e.g. Fidino et al., 2019; Hepler & Erhardt, 2021; Pautrel et al., 2024; Rota et al., 2016; Wohner et al., 2023). The complexity of multi-state and multi-species models, which require estimating additional parameters, likely accounts for this increased data demand. Conducting a traditional power analysis for site and survey numbers is essential to determine specific requirements for these types of models. Such an analysis should take into account the relationship between detection probability and occupancy probability, as both factors significantly influence the necessary scale of sites and surveys (Steenweg et al., 2019).

Case Study – What is possible with a low detection probability, but high occupancy dataset

The case study looking at different reproductive states of moose and roe deer in northern Sweden reveals insights into the challenges of using camera trap data for occupancy modelling, particularly when data is sparse and detection probabilities are low. The four-state models with daily detection histories failed to converge for both moose and roe deer, indicating that daily aggregation results in insufficient detection probabilities to inform the models adequately. This was evident by elevated $\hat{R}$ values, confidence intervals covering the entire parameter space, and low effective sample sizes. Weekly aggregation improved model fit, though certain parameters, particularly those involving calf presence in the moose model, remained difficult to estimate with confidence. This underscores the importance of sufficient detection probabilities, suggesting weekly aggregation as a more reliable basis for occupancy modelling, though it may still fall short for some parameters when data is sparse. These insights align with findings by Pautrel et al. (2024) and Steenweg et al. (2019), who found that while data aggregation can improve model fit and reduce uncertainty at low detection probabilities, it makes less to no difference in occupancy estimates when detection probabilities do not meet certain thresholds.

In contrast, the three-state models with both daily and weekly detection histories achieved convergence for all parameters, with acceptable confidence intervals for most, except for slightly larger intervals with weekly data. This indicates that reducing the complexity of the state space can enhance model performance even with lower detection probabilities. However, it also echoes Steenweg et al. (2019) which found that when detection probabilities are low, occupancy estimates are less accurate. Additionally, the failure of all models incorporating covariates to converge highlights a critical limitation: low detection probabilities severely constrain the ability to make inferences when adding covariates. This suggests that, despite aggregation strategies, the inflated number of model parameters when adding covariates may require fundamentally higher detection probabilities or larger sample sizes to achieve robust and reliable occupancy estimates in complex models. My preliminary simulations using 200 sites (Appendix III) confirm that simply increasing site numbers does not compensate for very low detection probabilities, emphasizing the need for careful study design and adequate data collection to ensure model reliability.

Strategies to increase detection probabilities in wildlife studies using camera traps can often introduce biases that compromise model assumptions (Hofmeester et al., 2019; MacKenzie et al., 2017). Common strategies include deploying multiple cameras per site (Evans et al., 2019; Hofmeester et al., 2021), using bait (e.g., salt licks, peanut butter, or sardines) (Sebastián-González et al., 2020), or positioning cameras in favourable spots like animal tracks (Burton et al., 2015). While deploying multiple cameras per site might not necessarily violate model assumptions, it is frequently impractical due to the limited availability of cameras in many study designs. Using bait or strategically placing cameras, although effective in increasing detection rates, often compromises the model's assumption of random sampling (Braczkowski et al., 2016; Wearn et al., 2013). However, these strategies may be required to collect adequate data for functional models, especially when studying evasive species or in areas with low density. The extent to which these biases affect the validity of the model is not entirely clear. However, when a grid-based study design is used and the bait only attracts animals from within the designated grid cell, the resulting data can still support robust models for questions on a larger scale, such as landscape and geographical habitat selection (Gilbert et al., 2021; Hofmeester et al., 2021). It is important to balance the need for higher detection probabilities while adhering to the model assumptions. When possible, employing multiple cameras per site is preferred to avoid bias, but other strategies may be justified if they significantly improve data quality and model functionality. Thus, these methods could provide valuable opportunities to enhance our understanding of wildlife ecology and inform conservation efforts.

This study underscores the importance of achieving high detection probabilities for reliable occupancy modelling, particularly in multi-state models and when incorporating covariates. Ensuring detection probabilities exceed 0.1, with values above 0.2 being preferable, is essential to minimize bias, reduce variability, and ensure model convergence. While data aggregation can alleviate some issues, it is not a replacement for adequate data collection. Our findings suggest that more than 60 sites are required for accurate inferences, especially when using covariates without strong prior knowledge. The case studies with moose and roe deer illustrate the challenges posed by low detection probabilities and demonstrate the benefits of weekly data aggregation. Researchers must balance the need for higher detection probabilities with adherence to model assumptions, employing strategies like multiple cameras per site where feasible. These insights provide guidelines for designing future wildlife studies using camera traps.

References

Bondi, N. D., White, J. G., Stevens, M., & Cooke, R. (2010). A comparison of the effectiveness of camera trapping and live trapping for sampling terrestrial small-mammal communities. *Wildlife Research*, *37*(6), 456–465. https://doi.org/10.1071/WR10046

Braczkowski, A. R., Balme, G. A., Dickman, A., Fattebert, J., Johnson, P., Dickerson, T., Macdonald, D. W., & Hunter, L. (2016). Scent lure effect on camera-trap based leopard density estimates. *PLOS ONE*, *11*(4), e0151033. https://doi.org/10.1371/journal.pone.0151033

Burton, A. C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J. T., Bayne, E., & Boutin, S. (2015). Wildlife camera trapping: A review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675–685. https://doi.org/10.1111/1365-2664.12432

Cromsigt, J., Ericsson, G., Dressel, S., Felton, A., Johansson, M., Neumann, W., Pfeffer, S., Sandström, C., Singh, N., Sjölander-Lindqvist, A., Spitzer, R., Widemo, F., & Widén, A. (2023). *Adaptiv sam- och flerartsförvaltning av klövvilt—Gemensam rapport för forskningsprojekten Beyond Moose och Governance* (p. 109). Naturvårdsverket. https://www.naturvardsverket.se/publikationer/7100/978-91-620-7108-0/

Delisle, Z. J., Flaherty, E. A., Nobbe, M. R., Wzientek, C. M., & Swihart, R. K. (2021). Next-generation camera trapping: Systematic review of historic trends suggests keys to expanded research applications in ecology and conservation. *Frontiers in Ecology and Evolution*, *9*. https://doi.org/10.3389/fevo.2021.617996

Delisle, Z. J., Henrich, M., Palencia, P., & Swihart, R. K. (2023). Reducing bias in density estimates for unmarked populations that exhibit reactive behaviour towards camera traps. *Methods in Ecology and Evolution*, *14*(12), 3100–3111. https://doi.org/10.1111/2041-210X.14247

Efford, M. G., & Dawson, D. K. (2012). Occupancy in continuous habitat. *ECOSPHERE*, *3*(4), 32. https://doi.org/10.1890/ES11-00308.1

Evans, B. E., Mosby, C. E., & Mortelliti, A. (2019). Assessing arrays of multiple trail cameras to detect North American mammals. *PLOS ONE*, *14*(6), e0217543. https://doi.org/10.1371/journal.pone.0217543

Fidino, M., Simonis, J. L., & Magle, S. B. (2019). A multistate dynamic occupancy model to estimate local colonization–extinction 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, J. T. (2023). Camera trapping in ecology: A new section for wildlife research. *Ecology and Evolution*, *13*(3), e9925. https://doi.org/10.1002/ece3.9925

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology*, *35*(1), 88–100. https://doi.org/10.1111/cobi.13517

Hepler, S. A., & Erhardt, R. J. (2021). A spatiotemporal model for multivariate occupancy data. *Environmetrics*, *32*(2), e2657. https://doi.org/10.1002/env.2657

Hofmeester, T. R., Cromsigt, J. P. G. M., Odden, J., Andrén, H., Kindberg, J., & Linnell, J. D. C. (2019). Framing pictures: A conceptual framework to identify and correct for biases in detection probability of camera traps enabling multi-species comparison. *Ecology and Evolution*, *9*(4), 2320–2336. https://doi.org/10.1002/ece3.4878

Hofmeester, T. R., Mos, J., & Zub, K. (2024). Comparing direct (live-trapping) and indirect (camera-trapping) approaches for estimating the abundance of weasels (Mustela nivalis). *Mammalian Biology*, *104*(2), 141–149. https://doi.org/10.1007/s42991-023-00394-z

Hofmeester, T. R., Thorsen, N. H., Cromsigt, J. P. G. M., Kindberg, J., Andrén, H., Linnell, J. D. C., & Odden, J. (2021). Effects of camera-trap placement and number on detection of members of a mammalian assemblage. *Ecosphere*, *12*(7), e03662. https://doi.org/10.1002/ecs2.3662

Hofmeester, T. R., Young, S., Juthberg, S., Singh, N. J., Widemo, F., Andrén, H., Linnell, J. D. C., & Cromsigt, J. P. G. M. (2020). Using by-catch data from wildlife surveys to quantify climatic parameters and timing of phenology for plants and animals using camera traps. *Remote Sensing in Ecology and Conservation*, *6*(2), 129–140. https://doi.org/10.1002/rse2.136

Kays, R., Arbogast, B. S., Baker-Whatton, M., Beirne, C., Boone, H. M., Bowler, M., Burneo, S. F., Cove, M. V., Ding, P., Espinosa, S., Gonçalves, A. L. S., Hansen, C. P., Jansen, P. A., Kolowski, J. M., Knowles, T. W., Lima, M. G. M., Millspaugh, J., McShea, W. J., Pacifici, K., … Spironello, W. R. (2020). An empirical evaluation of camera trap study design: How many, how long and when? *Methods in Ecology and Evolution*, *11*(6), 700–713. https://doi.org/10.1111/2041-210X.13370

Kellner, K. (2024). *jagsUI: A wrapper around "rjags" to streamline "JAGS" analyses. R package version 1.6.2* [Computer software]. https://kenkellner.com/jagsUI/.

Kéry, M., & Royle, J. A. (2020). *Applied hierarchical modeling in ecology: Analysis of distribution, abundance and species richness in R and BUGS* (Vol. 2, p. 783). Academic Press.

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: Second edition* (2nd ed., p. 641). Academic Press; Scopus. https://doi.org/10.1016/C2012-0-01164-7

Mackenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology*, *42*(6), 1105–1114. https://doi.org/10.1111/j.1365-2664.2005.01098.x

Mauget, C., Mauget, R., & Sempéré, A. (1997). Metabolic rate in female European roe deer (*Capreolus capreolus*): Incidence of reproduction. *Canadian Journal of Zoology*, *75*(5), 731–739. https://doi.org/10.1139/z97-094

Mauget, C., Mauget, R., & Sempéré, A. (1999). Energy expenditure in European roe deer fawns during the suckling period and its relationship with maternal reproductive cost. *Canadian Journal of Zoology*, *77*(3), 389–396. https://doi.org/10.1139/z98-230

McIntyre, T., Majelantle, T. L., Slip, D. J., & Harcourt, R. G. (2020). Quantifying imperfect camera-trap detection probabilities: Implications for density modelling. *Wildlife Research*, *47*(2), 177–185. https://doi.org/10.1071/WR19040

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. https://doi.org/10.1002/jwmg.433

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.

O'Connor, K. M., Nathan, L. R., Liberati, M. R., Tingley, M. W., Vokoun, J. C., & Rittenhouse, T. A. G. (2017). Camera trap arrays improve detection probability of wildlife: Investigating study design considerations using an empirical dataset. *PLOS ONE*, *12*(4), e0175684. https://doi.org/10.1371/journal.pone.0175684

Palmeirim, A. F., Benchimol, M., Peres, C. A., & Vieira, M. V. (2019). Moving forward on the sampling efficiency of neotropical small mammals: Insights from pitfall and camera trapping over traditional live trapping. *Mammal Research*, *64*(3), 445–454. https://doi.org/10.1007/s13364-019-00429-2

Pautrel, L., Moulherat, S., Gimenez, O., & Etienne, M.-P. (2024). Analysing biodiversity observation data collected in continuous time: Should we use discrete- or continuous-time occupancy models? *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.14314

Rich, L. N., Miller, D. A. W., Robinson, H. S., McNutt, J. W., & Kelly, M. J. (2016). Using camera trapping and hierarchical occupancy modelling to evaluate the spatial ecology of an African mammal community. *Journal of Applied Ecology*, *53*(4), 1225–1235. https://doi.org/10.1111/1365-2664.12650

Rota, C. T., Ferreira, M. A. R., Kays, R. W., Forrester, T. D., Kalies, E. L., McShea, W. J., Parsons, A. W., & Millspaugh, J. J. (2016). A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution*, *7*(10), 1164–1173. https://doi.org/10.1111/2041-210X.12587

Rovero, F., & Zimmerman, F. (2016). *Camera trapping for wildlife research*. Pelagic Publishing.

Sebastián-González, E., Morales-Reyes, Z., Naves-Alegre, L., Durá Alemañ, C. J., Gonçalves Lima, L., Machado Lima, L., & Sánchez-Zapata, J. A. (2020). Which bait should I use? Insights from a camera trap study in a highly diverse cerrado forest. *European Journal of Wildlife Research*, *66*(6), 99. https://doi.org/10.1007/s10344-020-01439-1

Severud, W. J., DelGiudice, G. D., & Obermoller, T. R. (2019). Association of moose parturition and post-parturition habitat with calf survival. *Journal of Wildlife Management*, *83*(1), 175–183. Scopus. https://doi.org/10.1002/jwmg.21570

Smith, P. A., Bart, J., Lanctot, R. B., McCaffery, B. J., & Brown, S. (2009). Probability of detection of nests and implications for survey design. *The Condor*, *111*(3), 414–423. https://doi.org/10.1525/cond.2009.090002

Solberg, E. J., Heim, M., Grøtan, V., Sæther, B.-E., & Garel, M. (2007). Annual variation in maternal age and calving date generate cohort effects in moose (Alces alces) body mass. *Oecologia*, *154*(2), 259–271. https://doi.org/10.1007/s00442-007-0833-9

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

Steenweg, R., Hebblewhite, M., Whittington, J., & McKelvey, K. (2019). Species-specific differences in detection and occupancy probabilities help drive ability to detect trends in occupancy. *Ecosphere*, *10*(4), e02639. https://doi.org/10.1002/ecs2.2639

Wearn, O. R., Rowcliffe, J. M., Carbone, C., Bernard, H., & Ewers, R. M. (2013). Assessing the Status of Wild Felids in a Highly-Disturbed Commercial Forest Reserve in Borneo and the Implications for Camera Trap Survey Design. *PLOS ONE*, *8*(11), e77598. https://doi.org/10.1371/journal.pone.0077598

Wevers, J., Beenaerts, N., Casaer, J., Zimmermann, F., Artois, T., & Fattebert, J. (2021). Modelling species distribution from camera trap by-catch using a scale-optimized occupancy approach. *Remote Sensing in Ecology and Conservation*, *7*(3), 534–549. https://doi.org/10.1002/rse2.207

Wohner, P. J., Scheerer, P. D., Meeuwig, M. H., & Peterson, J. T. (2023). A comprehensive multi-state conditional occupancy model for evaluating interactions of non-native and native species. *Frontiers in Ecology and Evolution*, *10*. https://doi.org/10.3389/fevo.2022.1014707

Acknowledgments

Firstly, I would like to thank Tim Hofmeester for wanting to supervise me, he has been a great inspiration and has helped me tremendously to keep me motivated throughout the thesis. I would also like to thank Henjo de Knegt for assisting me with conceptual modelling issues, study design, and running the largest part of the analysis on his power station. Without his help, this thesis would not have been of the same quality. Lastly, I would like to thank the students of the thesis room, who have made this time in Umeå amazing. It has helped me engage with my thesis on so many different levels and make it possible to help and receive help from everyone.

AI statement

Throughout this study, I utilized Open AI's Chat-GPT. It helped me assist in structuring R code as well as grammar and spelling mistakes. Chat-GPT was not employed for content creation in writing, but its role was strictly as a tool to solve errors or to increase the quality and readability of already written sentences. Furthermore, the responses provided were never integrated entirely, but only single words or snippets of good writing ideas were integrated into this thesis.

Appendix

Appendix I: Four State model with no covariates

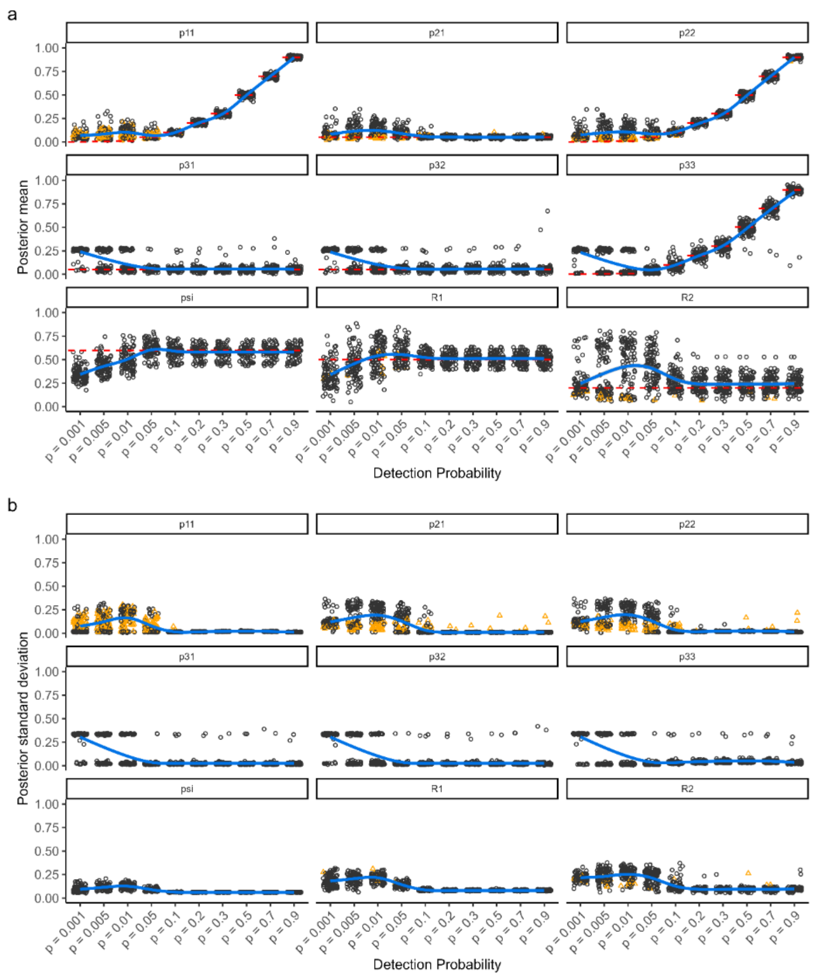

Figure 1: a) Mean of the posterior estimate for each fitted parameter (y-axis) plotted for 10 simulation scenarios with different detection probabilities (x-axis). b) Standard deviation of the simulation estimates for each fitted parameter value (y-axis), plotted for 10 different simulation scenarios with differing detection probabilities (x-axis). The blue lines indicate a GAM smooth function. The orange dots indicate that the model did not converge for this parameter having a rhat value >1.1. The red line indicates the true value. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

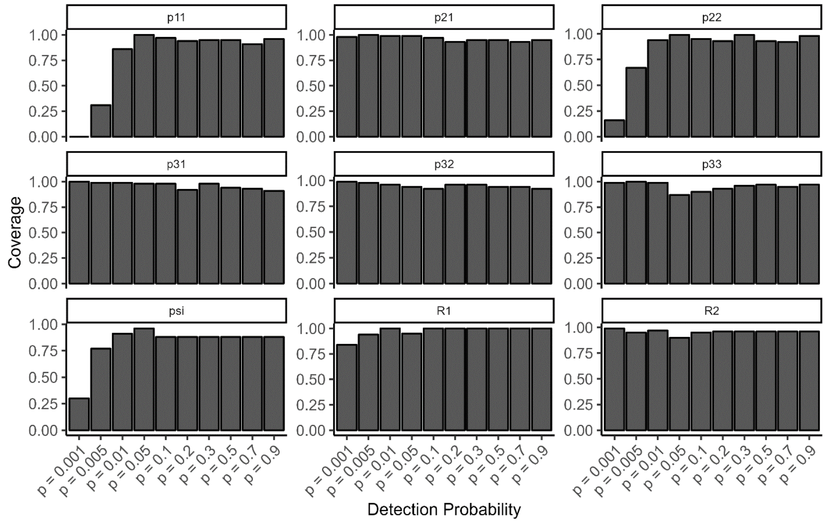

Figure 2: Coverage (y-axis) for each model parameter and simulated scenario (x-axis). The coverage is calculated using the posterior credible intervals derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

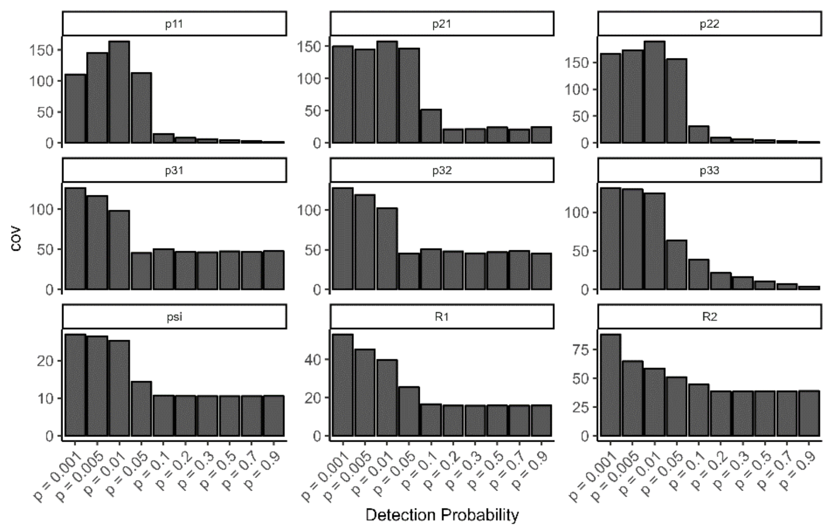

Figure 3: Coverage (y-axis) for each model parameter and simulated scenario (x-axis). The coefficient of variation is calculated using the posterior means and standard deviation derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

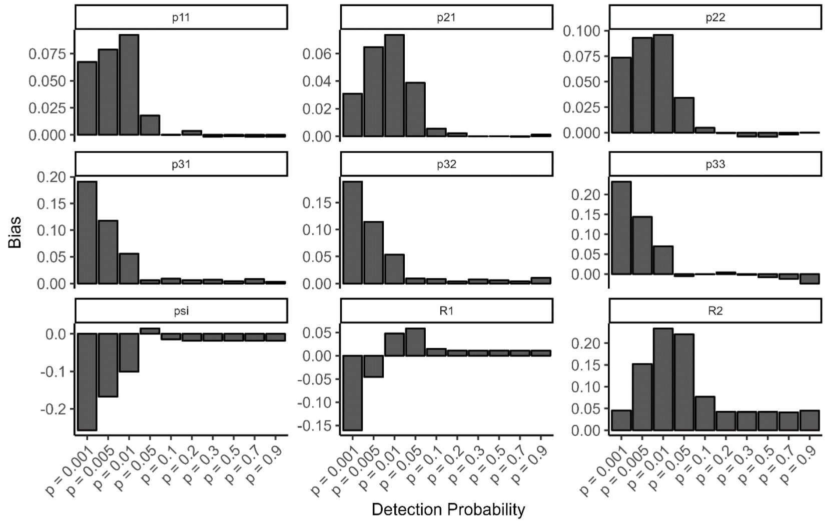

Figure 4: Bias (y-axis) for each model parameter and simulated scenario (x-axis). The bias is calculated using the posterior means and standard deviation derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

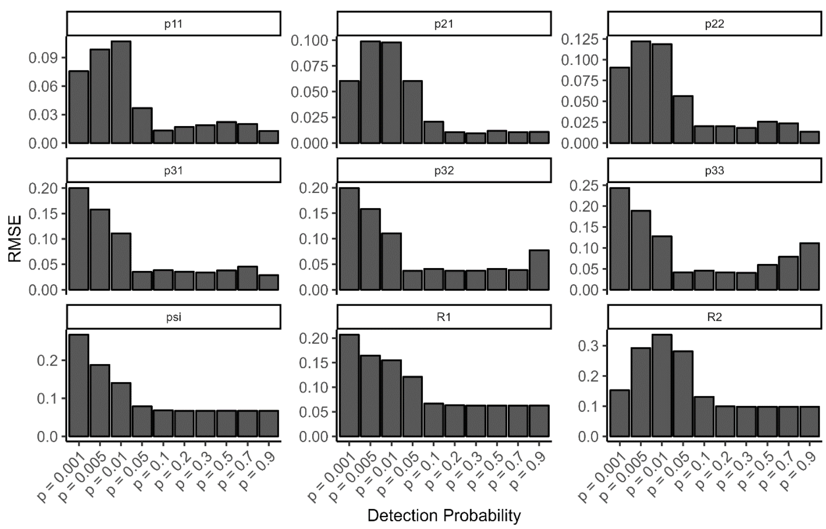

Figure 5: Root mean square error (y-axis) for each model parameter and simulated scenario (x-axis). The root mean quare error is calculated using the posterior means derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

Appendix II: Four State model with one covariate

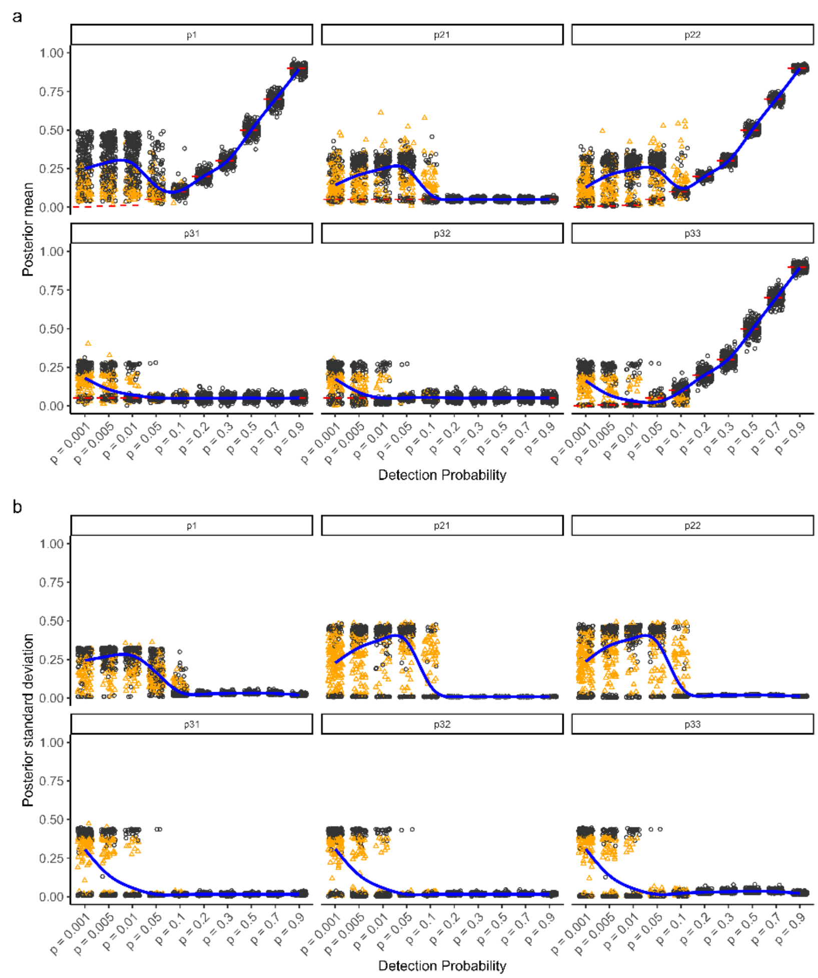

Figure 1: a) Mean of the posterior estimate for each fitted detection parameter (y-axis) plotted for 10 simulation scenarios with different detection probabilities (x-axis). b) Standard deviation of the simulation estimates for each fitted detection parameter value (y-axis), plotted for 10 different simulation scenarios with differing detection probabilities (x-axis). The blue lines indicate a GAM smooth function. The orange dots indicate that the model did not converge for this parameter having a rhat value >1.1. The red line indicates the true value. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

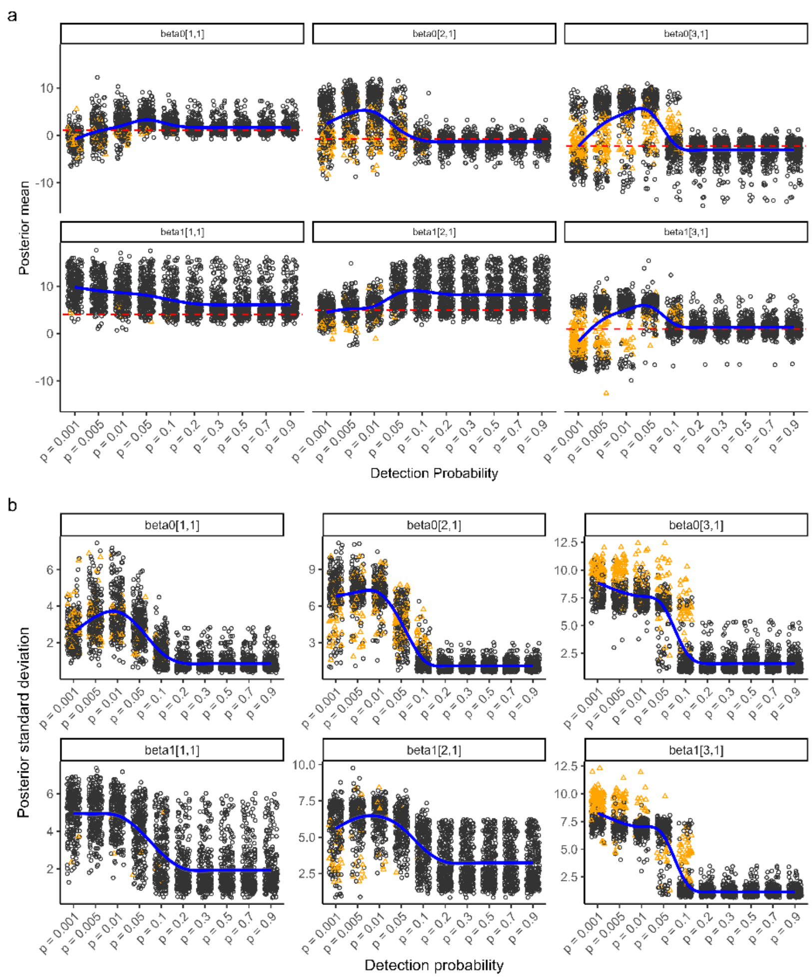

Figure 2: a) Mean of the posterior estimate for each fitted beta parameter (y-axis) plotted for 10 simulation scenarios with different detection probabilities (x-axis). b) Standard deviation of the simulation estimates for fitted beta parameter value (y-axis), plotted for 10 different simulation scenarios with differing detection probabilities (x-axis). The blue lines indicate a GAM smooth function. The orange dots indicate that the model did not converge for this parameter having a rhat value >1.1. The red line indicates the true value. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

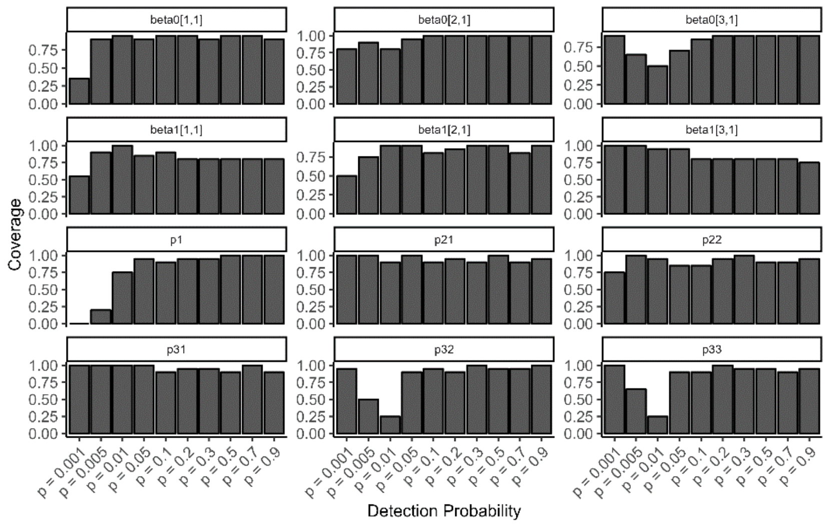

Figure 3: Coverage (y-axis) for each model parameter and simulated scenario (x-axis). The coverage is calculated using the posterior credible intervals derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

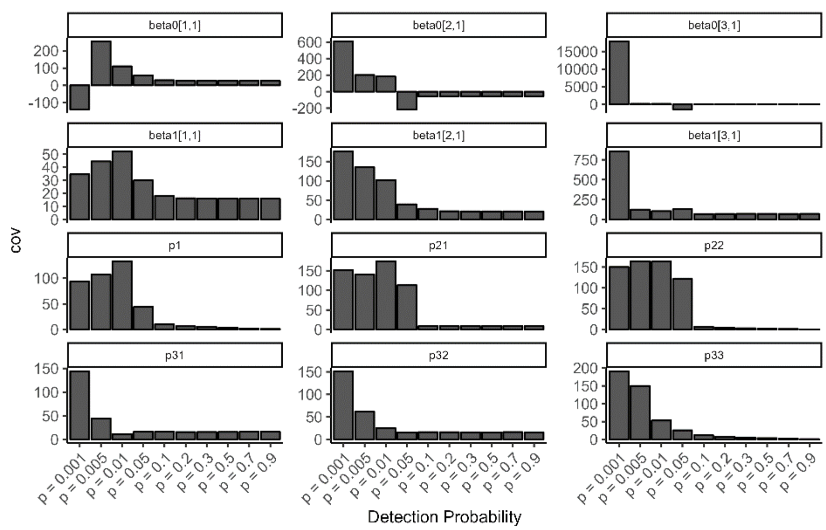

Figure 4: Coefficient of variation (y-axis) and simulated scenario (x-axis) for each model parameter. The coefficient of variation is calculated using the posterior means and standard deviation derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

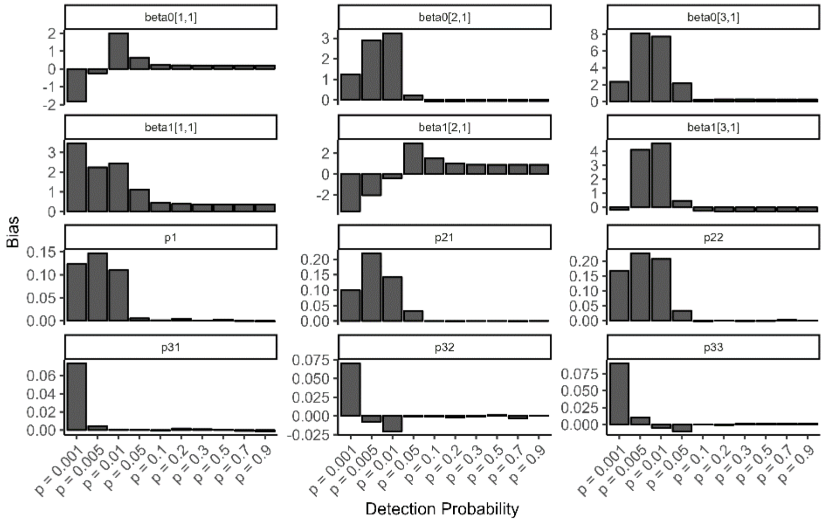

Figure 5: Bias (y-axis) and simulated scenario (x-axis) for each model parameter. The bias is calculated using the posterior means derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

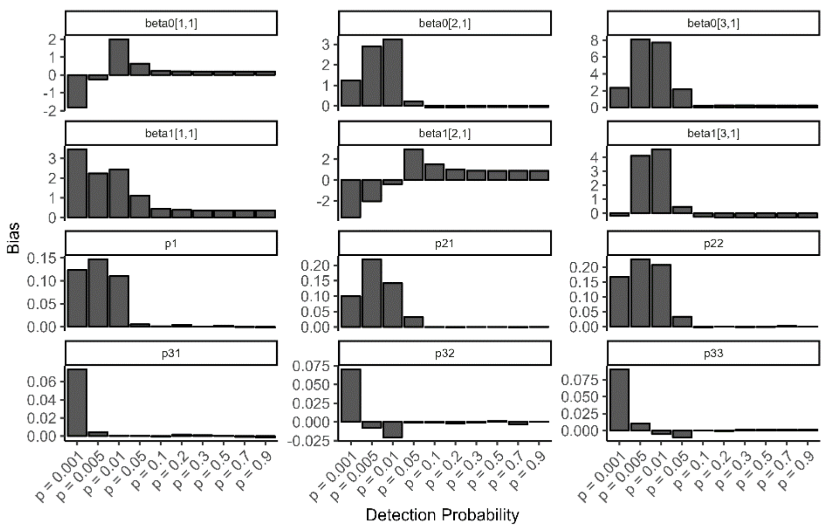

Figure 6: Root mean square error (y-axis) and simulated scenario (x-axis) for each model parameter. The errors are calculated using the posterior means derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

Appendix III: Four-state occupancy model simulated for 200 sites.

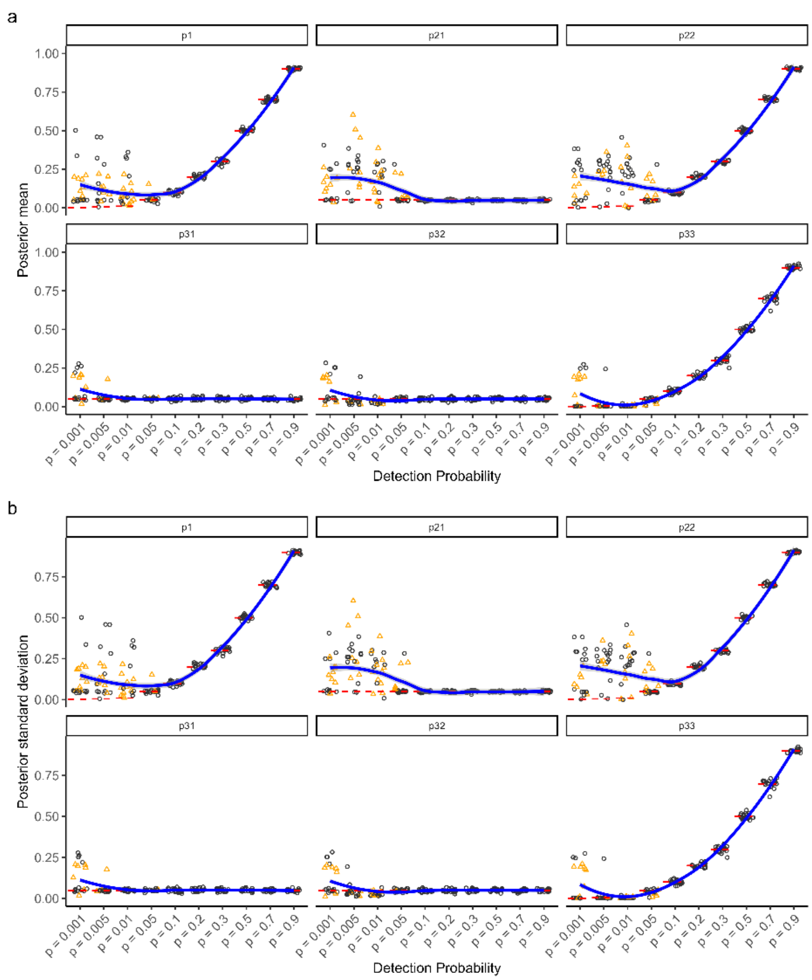

Figure 1: a) Mean of the posterior estimate for each fitted detection parameter (y-axis) plotted for 10 simulation scenarios with different detection probabilities (x-axis). b) Standard deviation of the simulation estimates for each fitted detection parameter value (y-axis), plotted for 10 different simulation scenarios with differing detection probabilities (x-axis). The blue lines indicate a GAM smooth function. The orange dots indicate that the model did not converge for this parameter having a rhat value >1.1. The red line indicates the true value. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

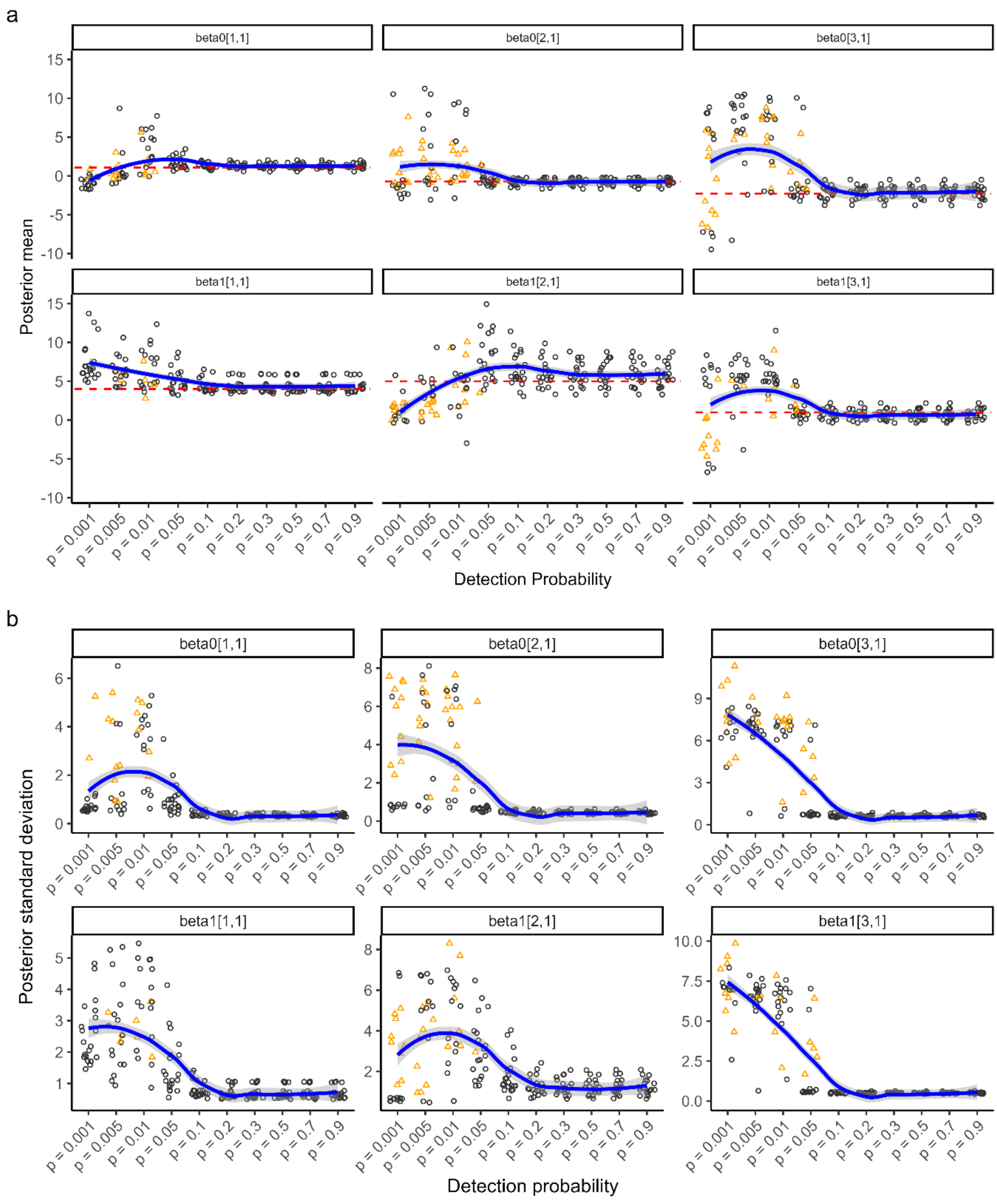

Figure 2: a) Mean of the posterior estimate for each fitted detection parameter (y-axis) plotted for 10 simulation scenarios with different detection probabilities (x-axis). b) Standard deviation of the simulation estimates for each fitted detection parameter value (y-axis), plotted for 10 different simulation scenarios with differing detection probabilities (x-axis). The blue lines indicate a GAM smooth function. The orange dots indicate that the model did not converge for this parameter having a rhat value >1.1. The red line indicates the true value. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

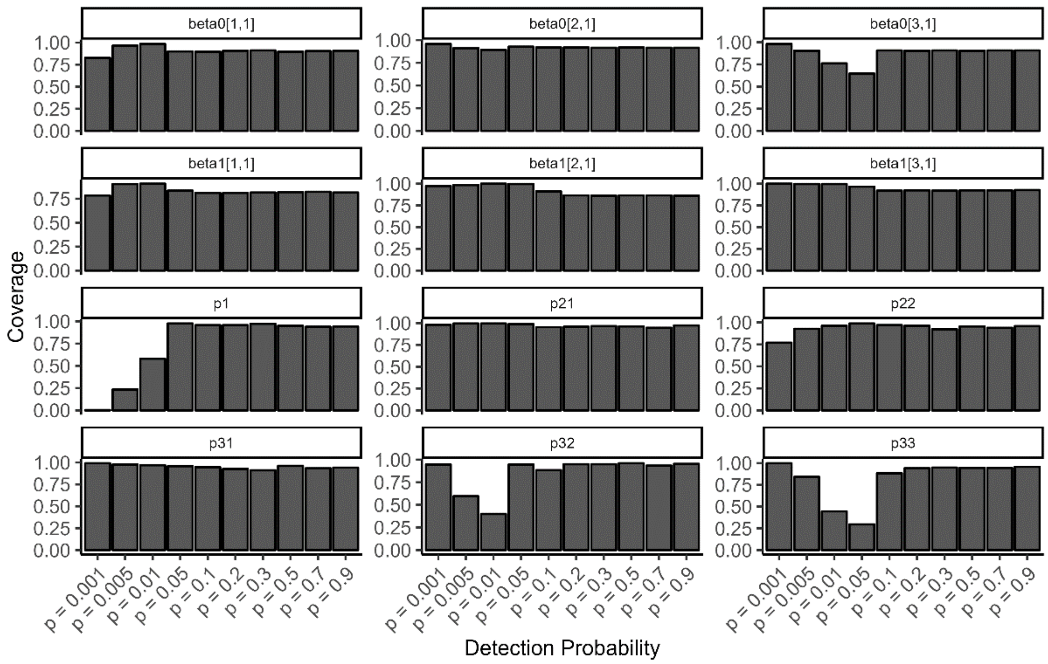

Figure 3: Coverage (y-axis) for each model parameter and simulated scenario (x-axis). The coverage is calculated using the posterior credible intervals derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

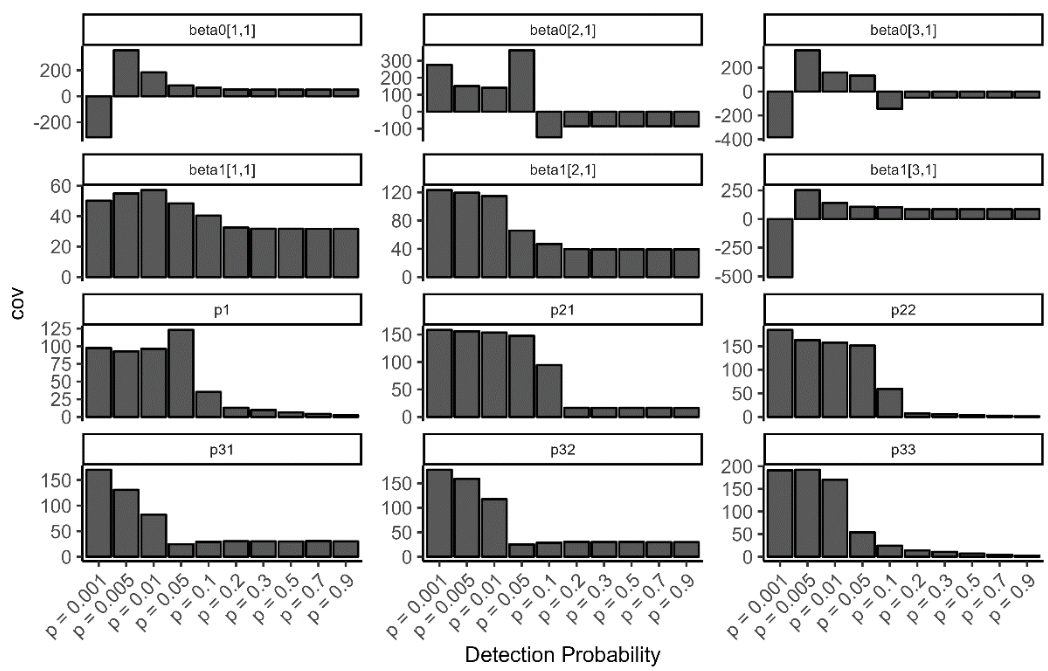

Figure 4: Coefficient of variation (y-axis) for each model parameter and simulated scenario (x-axis). The coefficient of variation is calculated using the posterior mean and standard deviation from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

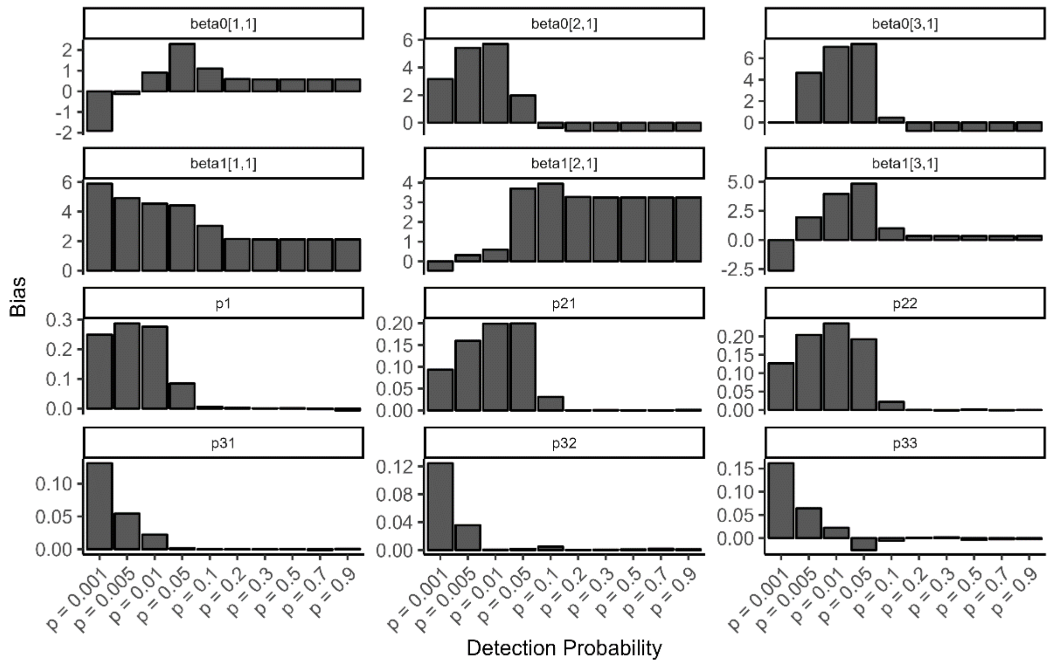

Figure 5: Bias (y-axis) for each model parameter and simulated scenario (x-axis). The bias is calculated using the posterior mean derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

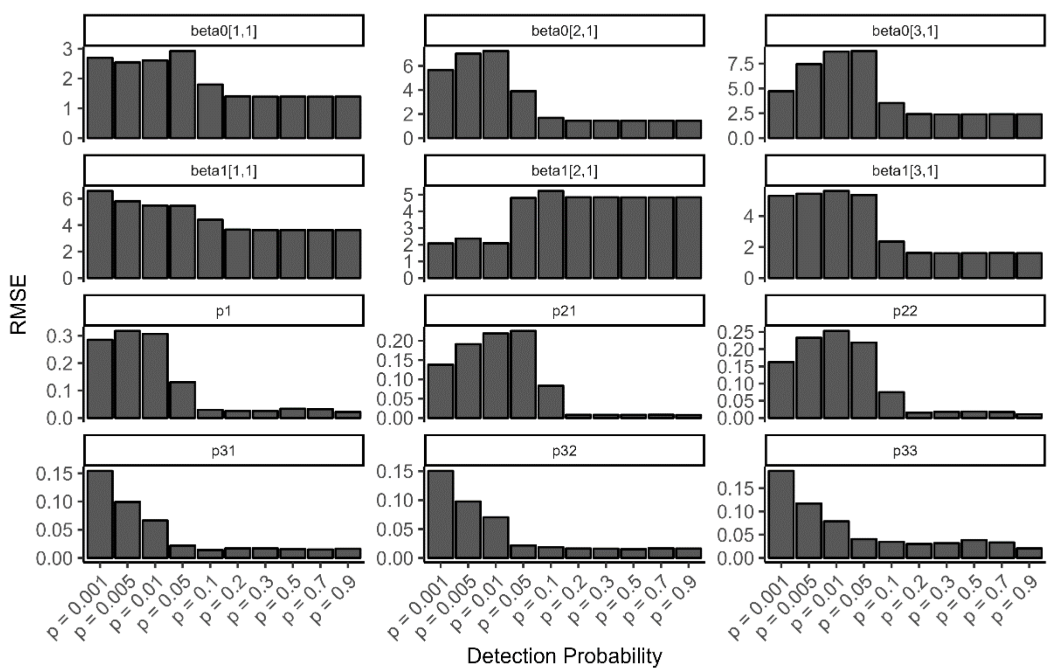

Figure 6: Root means square error (y-axis) for each model parameter and simulated scenario (x-axis). The bias is calculated using the posterior mean derived from the model simulations. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

Appendix IV: Case Study files with covariates

Four-states with covariates

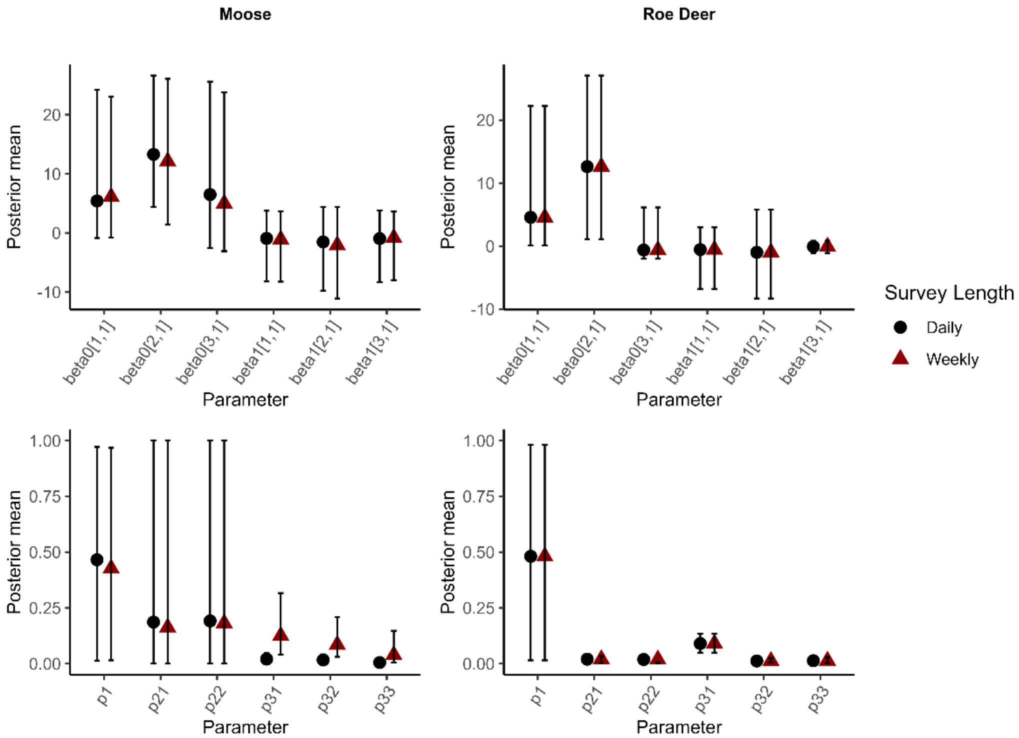

Figure 1: The posterior mean estimate and 95% credible intervals (y-axis) of both Moose and Roe deer three-state occupancy model. The detection and occupancy parameters are shown on the x-axis. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

Three-state with covariates

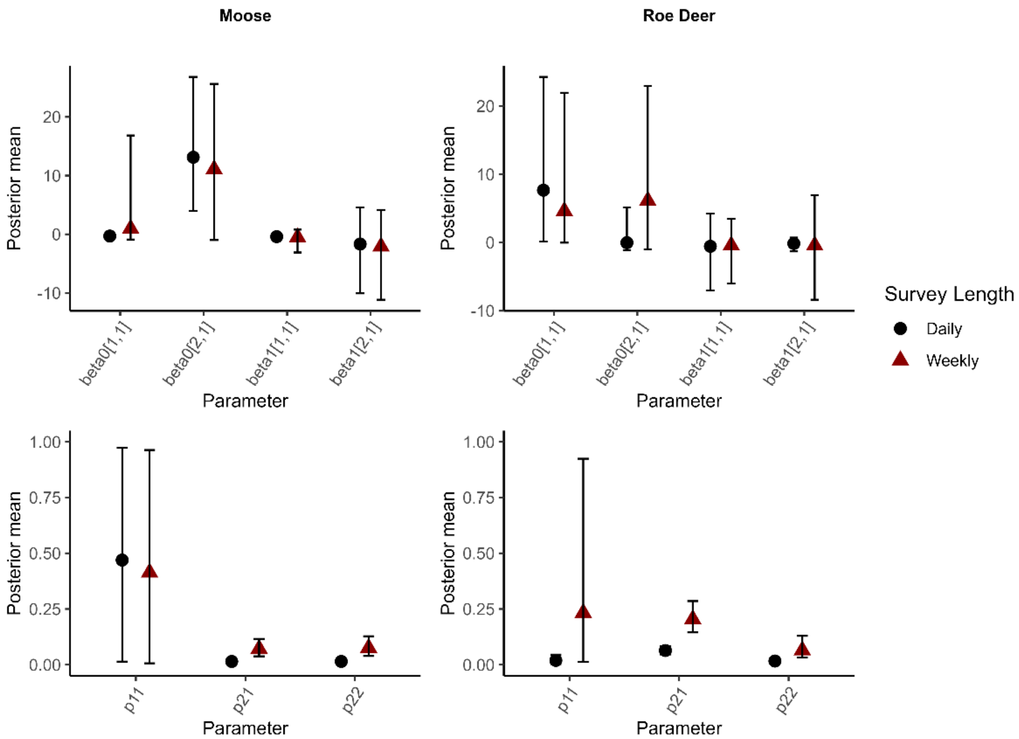

Figure 2: The posterior mean estimate and 95% credible intervals (y-axis) of both Moose and Roe deer three-state occupancy model. The detection and occupancy parameters are shown on the x-axis. See model description (chapter 2.1) for explanation of the detection and occupancy parameters.

Appendix V: Multi-state model with one covariate

```
model {
 # Model for the logit of the occupancy probabilities

  for (i in 1:nsites) {

    logit(psi[i]) <-   beta0[1,1]
                + beta1[1,1] * Fdens[i]
                 # Occupied

    logit(R1[i]) <-    beta0[2,1]
                + beta1[2,1] * Fdens[i]
                 # Occupied with 1 calf

    logit(R2[i]) <-    beta0[3,1]
                + beta1[3,1] * Fdens[i]
            # Occupied with multiple calves


  }

   # Priors for occupancy model coefficients - intercept and slope for forest (Need
another beta for each parameter, so if we want squared function )
   #so if we want squared function it will be beta0[i,k], beta1[i,k] where k is the
covariate
   #and i the state (so I will always be 1:3 whereas k will be 1:length(covariates)
for (k in 1:1) {
  for (i in 1:3){

    beta0[i,k] ~ dnorm(0, 0.01)
    beta1[i,k] ~ dnorm(0, 0.01)
  }
}

  # (2) Observation process
  # Linear models in observation process
  p1 ~ dunif(0, 1)       # Detection prob for sites with no calves
  lp21 ~ dnorm(0, 0.001) # Logit-transformed prob for detection as no calf, given 1
calf
```

```
  lp22 ~ dnorm(0, 0.001) # Logit-transformed prob for correct detection, given 1 calf
  lp31 ~ dnorm(0, 0.001) # Logit-transformed prob for detection as no calf, given >2 calves
  lp32 ~ dnorm(0, 0.001) # Logit-transformed prob for detection as 1 calf, given >2 calves
  lp33 ~ dnorm(0, 0.001) # Logit-transformed prob for correct detection, given >2 calves

  # Transforming logit probabilities to standard probabilities
  p21 <- exp(lp21) / (1 + exp(lp21) + exp(lp22))
  p22 <- exp(lp22) / (1 + exp(lp21) + exp(lp22))
  p31 <- exp(lp31) / (1 + exp(lp31) + exp(lp32) + exp(lp33))
  p32 <- exp(lp32) / (1 + exp(lp31) + exp(lp32) + exp(lp33))
  p33 <- exp(lp33) / (1 + exp(lp31) + exp(lp32) + exp(lp33))


  #  Occupancy matrix Omega
  for (i in 1:nsites) {
    Omega[i,1] <- 1 - psi[i]                          # Prob. of unoccupied
    Omega[i,2] <- psi[i] * (1 - (R1[i]))              # Prob. of occupancy w/o calves
    Omega[i,3] <- psi[i] * R1[i] * (1 - R2[i])        # Prob. of occupancy with 1 calf
    Omega[i,4] <- psi[i] * R1[i] * R2[i]              # Prob. of occupancy with >1 calves
  }

  # Define observation matrix (Theta)
  # Order of indices: true state, observed state
  Theta[1,1] <- 1
  Theta[1,2] <- 0
  Theta[1,3] <- 0
  Theta[1,4] <- 0
  Theta[2,1] <- 1 - p1
  Theta[2,2] <- p1
  Theta[2,3] <- 0
  Theta[2,4] <- 0
  Theta[3,1] <- 1 - (p21 + p22)
  Theta[3,2] <- p21
  Theta[3,3] <- p22
  Theta[3,4] <- 0
  Theta[4,1] <- 1 - (p31 + p32 + p33)
  Theta[4,2] <- p31
```

```
  Theta[4,3] <- p32
  Theta[4,4] <- p33

  # Initial latent state modeling
  for (i in 1:nsites) {
    z[i] ~ dcat(Omega[i,])
  }

  # Observation equation
  for (i in 1:nsites) {
    for (j in 1:nsurveys) {
      y[i,j] ~ dcat(Theta[z[i],])
    }
  }

  # Derived quantities
  for (i in 1:nsites) {
    occ1[i] <- equals(z[i], 1)
    occ2[i] <- equals(z[i], 2)
    occ3[i] <- equals(z[i], 3)
    occ4[i] <- equals(z[i], 4)
  }
  n.occ[1] <- sum(occ1[]) # Sites in state 1 (Unoccupied)
  n.occ[2] <- sum(occ2[]) # Sites in state 2 (Occupied w/o calves)
  n.occ[3] <- sum(occ3[]) # Sites in state 3 (Occupied with 1 calf)
  n.occ[4] <- sum(occ4[]) # Sites in state 4 (Occupied with >2 calves)

mean.psi <- mean(psi)
mean.R1 <- mean(R1)
mean.R2 <- mean(R2)
mean.Omega1 <- mean(Omega[,1])
mean.Omega2 <- mean(Omega[,2])
mean.Omega3 <- mean(Omega[,3])
mean.Omega4 <- mean(Omega[,4])
}
```

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.