FAST AND EFFICIENT APPROACHES TO LARGE-SCALE OCCUPANCY MODELS

Allan Ernest Clark

A thesis presented for the degree of
Doctor of Philosophy
in the Department of Statistical Sciences
University of Cape Town
October 2022

Supervisor: A/prof Altwegg

Acknowledgements

I would like to thank my supervisor, Res Altwegg, for taking me on so many years ago; for sending me on my first ISEC (Montepellier 2014) where I was introduced to the world of Statistical Ecology.

Thank you to the staff in the Department of Statistical Sciences for often encouraging me and the friendly advice during the process of completing my PhD (even though it's the third one). Thank you Prof Dunne! You were always willing to listen to me, to guide where needed and always made insightful comments related to work and life in general. You always seemed to think that I was cleverer than what I really was. I am sorry that I wasn't able to finish before your passing!

Thank you Sulaiman Salau and Reshma Kassanjee for your help and encouragement. You guys are good friends and I am happy you both are in my life.

I am thankful to the University of Cape Town's Emerging Researcher Programme for their guidance and support, the National Research Foundation of South Africa for funding this PhD through the Thuthuka grant programme as well as funding received through the University Capacity Development Program of the National Research Foundation of South Africa.

Certain computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: `hpc.uct.ac.za`.

Lastly, thank you to Anneke (and my family (My parents, Ouma, Jonathan, Peter-John, Mure, John and Anushka)) for the many sacrifices you gave up while I tried to complete my studies. Its been many years and I am glad that its over! Thank you for all the reading you had to do as well your constant support over the years. *Maybe now we can publish something together?*

Please note: an editor has not been used in the construction of this thesis.

Abstract

FAST AND EFFICIENT APPROACHES TO LARGE-SCALE OCCUPANCY MODELS

Allan Ernest Clark

6 October 2022

Bayesian occupancy models are important statistical tools that are used to investigate species range dynamics, species interactions as well as undercover key biological processes that drive occupancy (and detection) in a particular region. The results from these models are used to answer pressing conservation questions and are often used to develop responses to them. My original contribution to knowledge is the development of statistical methods that use detection-nondetection data commonly collected when undertaking occupancy modelling. The efficacy of these methods are investigated and applied to various South African detection-nondetection data sets.

I developed two Variational Bayes approximations to the posterior distribution of the parameters of a single season site occupancy model that uses logistic or probit link functions to model the probability of species occurrence at sites and species detection probabilities. The results suggest that under certain circumstances, the variational distributions provide accurate approximations to the true posterior distributions of the parameters of the model when the survey occasions are as low as three and the accuracy of the approximations improves as the sampling occasions increase. Approximate methods could be implemented when the detection probability is at least 0.5 and when there are at least three sampling occasions.

The link between logistic regression and occupancy modelling was exploited to develop a Gibbs sampler required to obtain posterior samples from the posterior distribution of the parameters of various occupancy model types (nonspatial, spatial and multi-species) when logit link functions are used to model the regression effects of the detection and occupancy processes. For the nonspatial occupancy model, the Gibbs sampling algorithm developed produces posterior samples that are identical to those obtained when using *JAGS* and *Stan* and that in certain cases the posterior chains mix faster than those obtained when using *JAGS*, stocc, and *Stan*. The Gibbs sampling algorithm developed for the multi-species occupancy model produces posterior samples that are identical to those obtained when using *Stan*, resulting in faster implementation times and a larger expected sampling rate than *Stan*.

The algorithms are implemented in the R package Rcppocc and MSO which is freely available on *GitHub* (`https://github.com/AllanClark/Rcppocc`, `https://github.com/AllanClark/MSO`).

Plagiarism declaration

This thesis has been submitted to Turnitin and I confirm that my supervisor has seen the report.

I hereby declare that all the work presented in this thesis, titled '**Fast and efficient approaches to large-scale occupancy models**' is my own work, except where otherwise stated in the text. This thesis has not been submitted in whole or in part for a degree at any other university.

Name	**Date**	**Signature**
Allan Ernest Clark (CLRALL001)	7 April 2022	

Declaration: inclusion of publications in the thesis

The statistical work in this thesis is my own! The work was carried out under the main supervision of Associate professor Res Altwegg while Associate professor John Ormerod was a co-author of Paper 1 below.

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publications in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the following publications:

- Clark A.E., Altwegg R, Ormerod JT (2016) A Variational Bayes Approach to the Analysis of Occupancy Models. PLoS ONE 11(2): e0148966. (Referred to as 'Paper 1'.)

- Clark, A.E. and Altwegg, R., 2019. Efficient Bayesian analysis of occupancy models with logit link functions. Ecology and Evolution, 9(2), pp.756-768. (Referred to as 'Paper 3'.)

The following chapters will be submitted for publication:

- Clark A.E., Altwegg R. A Variational Bayes approach to the analysis of occupancy models - the probit case. (Referred to as 'Paper 2'.)

- Clark A.E., Altwegg R. Gibbs sampling algorithms for multi-species site occupancy models. (Referred to as 'Paper 4'.)

Author contribution statement - Paper 1

Manuscript Reference

Clark A.E., Altwegg R, Ormerod JT (2016) A Variational Bayes Approach to the Analysis of Occupancy Models. PLoS ONE 11(2): e0148966. doi:10.1371/journal.pone.0148966.

Author contributions

'Allan Ernest Clark'[1,2] wrote the paper and performed the mathematical development and analysis. 'Res Altwegg'[1,2], 'John Ormerod'[3] and 'Allan Ernest Clark' reviewed the manuscript extensively. All authors have read and approved the manuscript.

Name	Date	Signature
Assoc prof Res Altwegg	16 March 2022	
Assoc prof John Ormerod	11/03/2022	
Allan Ernest Clark	27 March 2022	

Affiliations as at March 2022

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.
3. School of Mathematics and Statistics, University of Sydney, Sydney, Australia.

Author contribution statement - Paper 2 and 4

Comment

Chapter 3 and 5 of the thesis have not been published but are currently being compiled for future publication. The title of the proposed articles are 'A Variational Bayes approach to the analysis of occupancy models - the probit case' and 'Gibbs sampling algorithms for multi-species site occupancy models' respectively.

Author contributions

'Allan Ernest Clark'[1,2] wrote the paper and performed the mathematical development and analysis. 'Res Altwegg'[1,2] and 'Allan Ernest Clark' reviewed the manuscript extensively. All authors have read and approved the manuscript.

Name	Date	Signature
Assoc prof Res Altwegg	16 March 2022	
Allan Ernest Clark	27 March 2022	

Affiliations as at March 2022

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.

Author contribution statement - Paper 3

Manuscript Reference

Clark, A.E. and Altwegg, R., 2019. Efficient Bayesian analysis of occupancy models with logit link functions. Ecology and Evolution, 9(2), pp.756-768.

Author contributions

'Allan Ernest Clark'[1,2] wrote the paper and performed the mathematical development and analysis. 'Res Altwegg'[1,2] and 'Allan Ernest Clark' reviewed the manuscript extensively. All authors have read and approved the manuscript.

Name	Date	Signature
Assoc prof Res Altwegg	16 March 2022	
Allan Ernest Clark	27 March 2022	

Affiliations as at March 2022

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.

Conference presentation of chapters

Portions of the work in this thesis were presented at various conferences. These are listed below.

1. **Paper 1**
    - An oral presentation at
        - the *International Statistical Ecology conference*, Montpellier, France, 1-4 July 2014. I obtained the runner-up position in the PhD. student oral-competition.
2. **Paper 3**
    - Poster presentations at:
        - the *Euring 2017 Analytical Meeting*, Barcelona, Spain, 2-7 July 2017 and
        - the *International Society for Bayesian Analysis (ISBA) World Meeting*, Edinburgh, Scotland, 24-29 July 2018.
    - Oral presentations at:
        - the *International Statistical Ecology conference*, St Andrews, Scotland, 2-6 July 2018 and
        - *The South African Statistical Association conference*, Johannesburg, South Africa, 26-29 November 2018.
3. **Paper 4**
    - A poster presentation at:
        - *34*$^{\text{th}}$ *International Workshop on Statistical Modelling, Guimarães, Portugal, 7-12 July 2019.*
    - Oral presentations at:
        - the *International Statistical Ecology conference*, St Andrews, Scotland, 2-6 July 2018 and
        - *The South African Statistical Association conference*, Port Elizabeth, South Africa, 25-29 November 2019.

Contents

List of Figures

List of Tables

List of Abbreviations

AETPET	Ratio of potential to realised evapotranspiration
Anysberg	Anysberg Nature Reserve
dRUM	Difference of random utility method
ESR	Expected sampling rate
ESS	Expected sample size
G	Gaussian distribution
GM	Linear Gaussian mixture distribution
ICAR	Integrated conditional autoregressive (model or precision matrix)
INLA	Integrated nested Laplace approximation
ISEC	International Statistical Ecology Conference
JAGS	Just Another Gibbs Sampler
KL	Kullback-Leibler divergence
LOO	Approximate leave-one-out cross-validation
MCMC	Markov chain Monte Carlo
MH	Metropolis Hastings algorithm
MLE	Maximum likelihood estimator
MSAVI2	Modified soil-adjusted index
MSO	Multi-species occupancy (model)
nspp	Number of species observed by the birder
PAO	Proportion of sites occupied
PG	A Pólya-Gamma random variable
QDGC	Quarter-degree grid cell
RSR	Restricted spatial regression
SABAP2	2^{nd} Southern African Bird Atlas Project
SSO	Single-season occupancy (model)
VB	Variational Bayes
WAIC	Watanabe–Akaike information criterion
RJMCMC	Reversible jump MCMC

Chapter 1

Introduction

1.1 Study background

From a conservation point of view, we are often interested in whether a focal species occurs in a particular region; whether its' range is expanding, contracting, or remains the same, and thereafter understanding biological reasons for such observations. These questions and ones similar to those could be answered by making use of *species distribution models*. These models are statistical tools that utilise species occurrence or abundance data to obtain range maps of a species and display the probability that a species occurs in a geographical region. *Occupancy models* (MacKenzie et al., 2002) are a particular type of species distribution model whose outputs can be used to not only map the range distribution of species but can also be used to aid any conservation decision that relies on knowing whether a species occurs in a particular region or not and thus the advancement and study of these statistical models are of great value.

MacKenzie et al. (2002) formulated the single season occupancy model by using ideas borrowed from closed population mark-recapture models (Otis et al., 1978). In this model a number of sites (or regions) are visited (*surveyed*) a number of times to estimate the *occupancy* and *detection probabilities* of a species associated with each *site*. Here the term *occupancy probability* refers to the probability that a particular site is inhabited by a species whereas the term *detection probability* refers to the probability of detecting the species during a survey given that the species occupies the site. Here, the term *detection* refers to whether or not the species was observed during a particular survey of the site when they are present, and lastly, a *site* is viewed as a spatial unit on which occupancy is defined. These important introductory concepts are further discussed in Section 1.1.2 below.

Occupancy models are an improvement on previous methods (generalized linear models (McCullagh and Nelder, 1989) or generalized additive models (Hastie and Tibshirani, 1990) since they explicitly account for the potential non-detection of species at surveyed sites. Failure to account for false negatives (non-detections when the site is truly occupied) often leads to biased occupancy probabilities (Guillera-Arroita et al., 2014) with the potential of making incorrect management decisions regarding the conservation of animal species.

A number of different types of occupancy models have been developed since 2002. A non-exhaustive list of model types includes the following: (i) single season occupancy model (MacKenzie et al., 2002), (ii) multi-season (or dynamic) occupancy model (MacKenzie et al., 2003; Royle and Kery, 2007), (iii) multi-state occupancy models (MacKenzie et al., 2009; Martin et al., 2009), (iv) models that account for false-positive detections (Royle and Link, 2006; McClintock et al., 2010; Miller et al., 2011) and (v) multi-species occupancy models (Dorazio and Royle, 2005; Tobler et al., 2015; Broms et al., 2016). Section 1.1.2 briefly introduces the relevant models discussed in this thesis.

The occurrence of a species tends to be correlated across space because individuals move and because nearby places tend to have similar environments. Early developments of occupancy models neglected to account for such spatial autocorrelation in the occupancy process. Models including conditional autoregressive models, improper conditional autoregressive models and more recently restricted spatial regression models which are more complex than the initial nonspatial models (Gelfand et al., 2005; Johnson et al., 2013; Clark and Altwegg, 2019) have been developed to account for spatial autocorrelation in the occupancy process.

As the complexity of different occupancy models increased (e.g., the inclusion of random effects in the occupancy and detection processes), a Bayesian approach to occupancy modelling has been preferred to the maximum likelihood approach. Software such as *WinBUGS* (Lunn et al., 2000), *JAGS* (Plummer, 2003), *OpenBUGS* (Spiegelhalter et al., 2007) and more recently *Stan* (Carpenter et al., 2017) and *NIMBLE* (de Valpine et al., 2017) are often used to run such analyses. Such programming platforms allow for very general and complicated model structures to be easily incorporated at the expense of added computational times since Bayesian models utilise Markov chain Monte Carlo (MCMC) methods (Brooks et al., 2011) and are thus more computationally intensive when compared to maximum likelihood estimation. In an occupancy modelling context, Joseph (2012) demonstrates how to reduce computational run-times by marginalising across the latent discrete (occupancy) state. Yackulic et al. (2020) further investigates this idea and applies it to a number of different statistical models that incorporates latent discrete states. Run-time comparisons were made between *WinBUGS*, *JAGS*, *OpenBUGS* and *Stan* and they found that *Stan* significantly outperformed other programming platforms. Turek et al. (2017) noted that the optimal MCMC sampling strategy to follow when analysing data is 'problem-specific' and thus there is no single strategy that works best when using MCMC algorithms (Ponisio et al., 2020).

Fast approximate methods named INLA (integrated nested Laplace approximations) have been developed to approximate Bayesian statistical models which have a continuous state space (Rue et al., 2009). Unfortunately since occupancy models have discrete

state spaces, INLA cannot be utilised to fit occupancy models (as formulated in Section 1.1.2). Variational Bayes is a method that approximates the posterior distribution of the parameters of a model by minimising the Kullback-Leibler distance (Kullback and Leibler, 1951) between the unknown posterior distribution and an approximating distribution. Few assumptions are made regarding the approximating distribution. Variational Bayes is often used as an alternative to MCMC since the methods are often much faster than MCMC without much loss in accuracy (Ormerod and Wand, 2010) and thus could potentially be used to fit occupancy models to data.

Occupancy models have successfully been used to predict species distributions as well as identify key habitat characteristics associated with species occurrence elsewhere (see Kéry et al. (2010), Peterman et al. (2013), Roth et al. (2014) to list but a few examples). In a South African context, Altwegg et al. (2008), investigated how the range dynamics of the Hadeda Ibis was affected by changes in climate by relating occupancy and detection probabilities to climatic as well as biome types. Bled et al. (2013) extended this work by building a complicated spatial occupancy model to investigate the range expansion of the Hadeda Ibis during the 1st and the 2nd South African bird Atlas (see Section 1.1.1 for a brief overview of this project). Broms et al. (2014) also used a spatial model to investigate how the range distribution of the Southern Ground Hornbill is affected by protected areas. Because of the large spatial scale and the need to model the *detection process* (see Section 1.1.2) in a sophisticated way, Bled et al. (2013) as well as Broms et al. (2014) used Bayesian methods (Royle and Dorazio, 2008) to fit their models. A big limitation, however, is that the current state of the art implementation of such spatial occupancy models is relatively slow and this limits the number of species that can be analysed using these methods. The current methods are slow since the estimation methods are not easily parallelised to reduce implementation times (Wilkinson, 2005). Also, the underlying coefficients are often highly correlated which leads to longer implementation times (Broms et al., 2014). The slow implementation times have prompted the use of non-Bayesian methods of identifying suitable covariates that could be included in modelling (Broms et al., 2014).

1.1.1 The 2nd Southern African Bird Atlas Project (SABAP2)

Three of the four chapters in this thesis use data extracted from the 2nd Southern African Bird Atlas Project (SABAP2). The SABAP2 project was established in 2007 to undertake various analyses and '... aims to map the distribution and relative abundance of birds in southern Africa and includes: South Africa, Lesotho, Botswana, Namibia, Mozambique, eSwatini, Zimbabwe, and Zambia' (Animal Demography Unit, 2014). The data consists of checklists collected by citizen scientists that spend time searching for different bird species and then submitting checklists to the database. Citizen scientists are required to make a list of all the species as well as the order in which they observed the species during at

least two hours of intense birding. They are allowed to add additional species to their list for up to five days and are additionally instructed to search in all habitat types within a sampling unit.

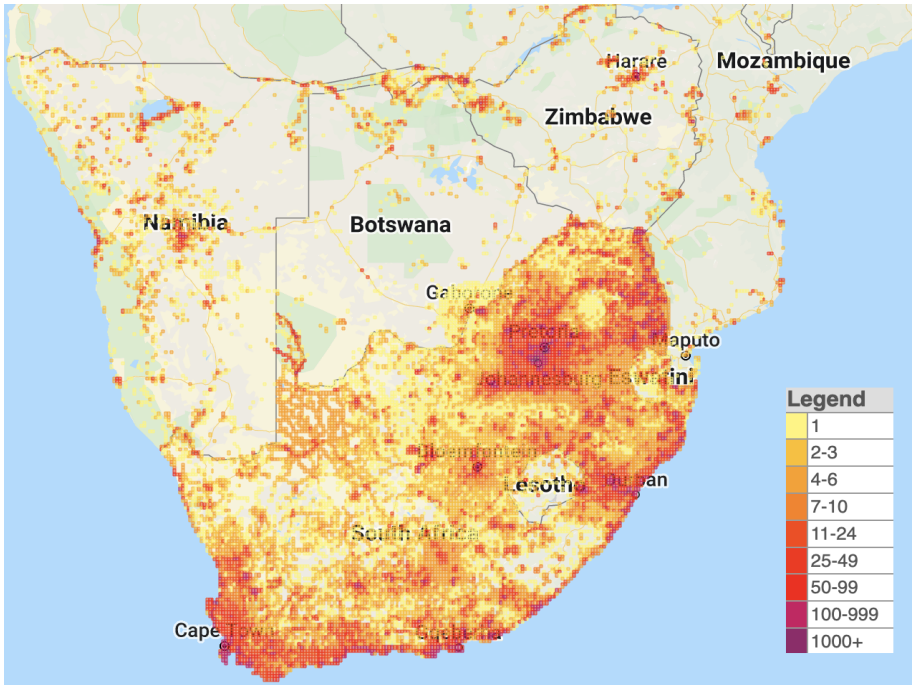

Figure 1. The spatial coverage of the SABAP2 data, as at 3 March 2022. The colour scale represents the number of checklists received per pentad.

The sampling unit of the project is a five-minute longitude by five-minute latitude grid cell (i.e. $5' \times 5'$), and is termed a *pentad*. SABAP2 consists of 17 444 pentads within South Africa and as at March 2022 over 17 million records have been collected. SABAP2 is a successor of SABAP1 that used $15' \times 15'$ grid cells termed a *quarter-degree grid cell* (QDGC). I did not use SABAP1 data. Nine pentads lie within each QDGC. Figure 1 displays the spatial coverage of the SABAP2 database. The colour scale used refers to the number of checklists per pentad and highlights the sampling coverage bias in the database (Hugo and Altwegg, 2017). In this thesis, I use pentads as sites and checklists as surveys (producing a detection or non-detection) in the occupancy models developed in Chapters Two-Four.

A very brief introduction to the types of occupancy modelling discussed in the thesis is provided next.

1.1.2 Single-season occupancy models

In what follows, assume that we are interested in investigating the occurrence of one particular species at a set of sites. Here the term *species* refers to any particular animal species, taxon, bird species, plant species, etc. In order to explain the relevant concepts

related to an occupancy analysis, consider Figure 2 (a) where a species occupies all **black** grid cells (used interchangeably as *sites*) and does not occupy **white** grid cells. The process that brings about the occupancy status of sites is known as the *ecological/occupancy process*. i.e. the process that brings about the black and white configuration of grid cells.

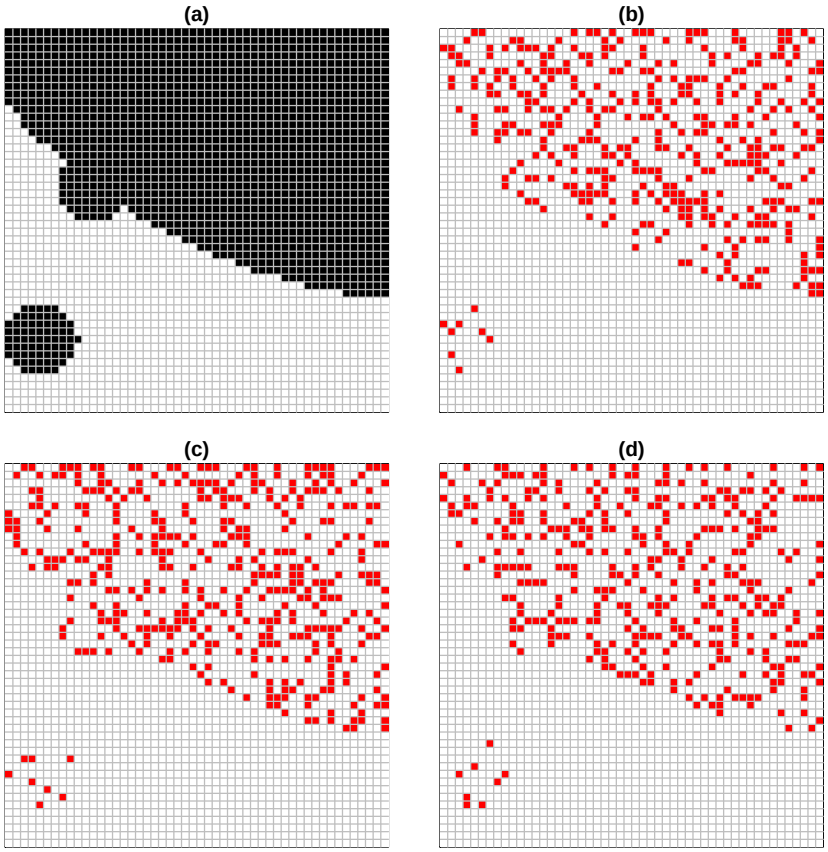

Figure 2. (a) An example occupancy process where a **black** grid cell represents an occupied site while a white grid cell is an unoccupied site; (b)-(d) are examples of survey data sets such that red grid cells represents sites where the species was detected while white grid cells represents sites where the species was not detected.

In the single season occupancy model (MacKenzie et al., 2002) each site is surveyed a number of times to estimate the *occupancy* - and *detection* - *probabilities* of a species associated with each site. The term *occupancy probability* refers to the probability that a particular site is *occupied* by a species, the term *detection probability* refers to the probability of detecting the species during a survey given that the species occupies the site while the mechanism governing whether or not a species is seen at a site is known as the *detection process*.

The data used to estimate the parameters of an occupancy model are stored in a matrix that contains zeros and ones which represents whether or not the species was observed during surveys for all sampled sites. Figures 2 (b)-(d) are examples of fictitious survey data for the same region displayed in Figure 2 (a). Red grid cells indicate that the species

was observed in a particular grid cell while a white grid cell indicates that the species was not detected in a grid cell. These figures display that even though the species occupies a particular site, it is not always observed during surveys of the region.

The three main assumptions underlying the statistical development of the single season occupancy model are (i) geographic and demographic closure of sites under investigation, (ii) independence and (iii) no false-identification of species. The *closure* assumption requires that sites are either occupied or not for the entire duration of the study. The *independence* assumption states that (i) the detection and occupancy probability of a site is independent of the occupancy and detection probability of another site; (ii) detections at a single site are independent of each other(Devarajan et al., 2020).

Using the above example as background, I briefly discuss the statistical formulation of the single season occupancy model. MacKenzie et al. (2002) considered a two parameter model where the occupancy (ψ) and detection (p) probabilities are constant for all sites and survey visits. Assume that detection-nondetection data is collected in an $n \times k$ matrix ($\mathbf{y}$) such that the number of surveys (k) are the same for all sites (n in total). In this case, the likelihood function is

$$\mathcal{L}(\psi, p|\boldsymbol{y}) = \prod_{i=1}^{n} \left(\psi p^{d_i} (1-p)^{k-d_i} + (1-\psi) I_{(d_i=0)} \right), \tag{1.1}$$

where $d_i = \sum_{j=1}^{k} y_{i,j}$ represents the number of detections at site i and $I_{(\text{condition})}$ is an indicator function that takes on the value 1 if the *condition* is satisfied and 0 otherwise.

The above formulation was extended to incorporate covariates to model site specific occupancy probabilities (ψ_i) and survey specific detection probabilities ($p_{i,j}$). The likelihood function in this case is

$$\mathcal{L}(\boldsymbol{\psi}, \boldsymbol{p}|\boldsymbol{y}) = \prod_{i=1}^{n} \left(\psi_i \left(\prod_{j=1}^{k} p_{i,j}^{y_{i,j}} (1-p_{i,j})^{1-y_{i,j}} \right) + (1-\psi_i) I_{(d_i=0)} \right), \tag{1.2}$$

where *logistic* functions are used to model ψ_i and $p_{i,j}$. The formulations of the occupancy model in equations (1.1) and (1.2) do not explicitly incorporate $\boldsymbol{z}$, the true occupancy status at each site. Information regarding $\boldsymbol{z}$ can be obtained by representing the above models using a hierarchical Bayesian model as follows (Royle and Dorazio, 2008)

$$z_i|\psi_i \sim \text{Bernoulli}(\psi_i) \text{ and} \tag{1.3}$$

$$y_{i,j}|z_i, p_{i,j} \sim \text{Bernoulli}(z_i p_{i,j}) \tag{1.4}$$

for all sites $i = 1, \ldots, n$; for all visits $j = 1, \ldots, V_i$ (here I assume that the number of visits to each site might not all be the same). The true occupancy status of a site (z_i) is a partially observed latent variable and takes on the value of 1 if a site is occupied and 0

otherwise.

Royle and Dorazio (2008) modelled the occupancy and detection processes using *logistic* link functions and used *WinBUGS* to fit their models. Shortly after this, Dorazio and Rodriguez (2012) utilised the connection between the *probit* link function and data augmentation (Albert and Chib, 1993) and developed a Gibbs sampling algorithm to fit a single season occupancy model. They stated that the choice of the link function used is subjective and that similar results would be obtained for different link functions if the occupancy and detection probabilities were not close to zero or one. Note however that, one of the main advantages to preferring a *logistic* link function is that it allows one to easily interpret the coefficient effects of an estimated model by using odds ratios, which is not easily done for a *probit* link function.

Due to the independence assumption, the above model formulation (equations (1.3) and (1.4)) does not account for spatial autocorrelation in either the detection or the occupancy processes. This shortcoming can easily be accounted for by modelling the occupancy and detection probabilities as functions of $f(.)$ and $g(.)$ as follows (assuming a *logistic* link function is used)

$$\begin{aligned} logistic(\psi_i) &= \boldsymbol{x}_i^T\boldsymbol{\beta} + f(.), \\ logistic(p_{i,j}) &= \boldsymbol{w}_{i,j}^T\boldsymbol{\alpha} + g(.) \end{aligned}$$

for all sites $i = 1, \ldots, n$; for all visits $j = 1, \ldots, V_i$. The $\boldsymbol{x}$ and $\boldsymbol{w}$ row-vectors are appropriately dimensioned and contain occupancy and detection covariates, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are regression effects while $f(.)$ and $g(.)$ are functions introduced to capture potential spatial autocorrelation in the occupancy and detection processes respectively.

Johnson et al. (2013) extended the single season occupancy model (specifically equation (1.3)) to account for spatial autocorrelation that might be present in the occupancy process by using spatially structured and unstructured random effects. They specifically modelled the occupancy process as a function of fixed (environmental) effects as well as spatial random effects ($f(.)$). Wright et al. (2019) used similar spatial effects when modelling the detection process while Mohankumar and Hefley (2021) recently suggested using machine learning methods to approximate the spatial structure in the occupancy process.

The models discussed in this section all pertain to modelling the occupancy and detection process of a single species (during a single season). In the following section I briefly discuss similar methods for modelling the group dynamics of multiple species. This is done by considering the multi-species occupancy model.

1.1.3 Multi-species occupancy model

Multi-species occupancy models use detection-nondetection data from species observed at different locations to estimate the probability that a species occupies a geographical region. The models can be used to estimate the occupancy probabilities associated with species that have too little data for single species models to be reliable (Devarajan et al., 2020) and are also used to obtain estimates of species richness (defined as the number of distinct species in a region) in a geographical region (Zipkin et al., 2010; Sauer et al., 2013). In certain geographical regions where species richness is known or well-understood (e.g. in well-studied areas) these models are also used to uncover the biological drivers of biodiversity in a region (Zipkin et al., 2009; Ruiz-Gutiérrez et al., 2010; Drouilly et al., 2018). Below I briefly provide the formulation of the multi-species occupancy when the aim of an analysis is to estimate the species richness in a region[1].

Dorazio and Royle (2005) developed a model that could be used to estimate the species richness (denoted as N) in a geographical region. They assume that there exists a super-population of species (denoted as S) that consists of the observed species, n_s, as well as additional unseen species, $S - n_s$. Here all information relating to survey occasions are grouped together such that the observed data $\boldsymbol{Y}_{n_s}$ consists of a $n_s \times J$ matrix of the total number of detections for each species at each location (from K_j surveys, $j = 1, \ldots, J$)[2]. Since N is unknown, an $(S - n_s) \times J$ matrix of zeros are introduced (a data augmentation step) which represent the data associated with the unobserved species.

A latent indicator variable w_i is introduced which takes on the value 1 if species i in the super-population occurs in the geographical region under investigation and 0 otherwise. From the above discussion, $w_i = 1$ for $i = 1, \ldots, n_s$ and the species richness in a geographical region is $N = \sum_{i=1}^{S} w_i$. The latent variable $z_{i,j}$ takes on the value 1 if species i occupies location j and 0 otherwise. The w_i indicator is modelled using a Bernoulli distribution with success parameter Ω while, $z_{i,j}|w_i$ is modelled as a Bernoulli random variable with success probability, $w_i\psi_{i,j}$. The detection process (conditional on $w_i = 1, z_{i,j} = 1$) is modelled using a binomial distribution such that

$$p(y_{i,j}|z_{i,j} = 1, p_{i,j}, w_i = 1) = \binom{K_j}{y_{i,j}} p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{K_j - y_{i,j}}.$$

Using the above description it is clear that the model can be formulated hierarchically as

$$w_i \sim \text{Bernoulli}(\Omega), \ \forall i = 1, \ldots, S$$

[1]Note that here I have not included any details with regards to the specification of the prior distributions of the parameters or hyper-parameters of the following model. These details can be found in Chapter 5.

[2]This is undertaken so that the detection process is modelled as a Binomial distribution. This simplifies the analysis somewhat and aids in the computational efficiency of the resulting estimation algorithm.

$$
\begin{aligned}
z_{i,j}|w_i, \psi_{i,j} &\sim \text{Bernoulli}(w_i\psi_{i,j}),\ \forall i = 1, \ldots, S;\ \ \forall j = 1, \ldots, J, \\
y_{i,j}|z_{i,j}, w_i, p_{i,j} &\sim \text{Binomial}(K_j,\ w_i z_{i,j} p_{i,j}), \\
p_{i,j} &= \Phi(\boldsymbol{v}_j^T \boldsymbol{\alpha}_i) \text{ with } \boldsymbol{\alpha}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}} \boldsymbol{I}_{n_d}), \\
\psi_{i,j} &= \Phi(\boldsymbol{x}_j^T \boldsymbol{\beta}_i) \text{ with } \boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}} \boldsymbol{I}_{n_o}),
\end{aligned}
$$

where $\boldsymbol{v}_j$ and $\boldsymbol{x}_j$ variables are site-specific covariates used to model the detection and occupancy processes respectively. When species richness is known, w_i is not required and the data augmentation step is not undertaken.

In the following section I discuss the research aims and objectives of the thesis.

1.2 Research aims and objectives

1.2.1 Aims

The main aim of this study is to develop statistical methods that can be used to fit various types of occupancy models to detection-nondetection data. The three main areas investigated are: (1) Variational Bayes methods applied to single season occupancy models; (2) the development of a fast and efficient MCMC algorithm for fitting Bayesian nonspatial and spatial single season occupancy models when the regression effects of the occupancy model are modelled using *logistic* link functions and (3) the development of various Gibbs sampling algorithms for fitting multi-species occupancy models. A secondary aim of the study is to develop a publicly available R package which allows users to implement the methods developed here on their data.

1.2.2 Objectives

The objectives of the thesis are as follows:

1. Derive Variational Bayes methods to approximate the posterior distributions of the parameters of a nonspatial occupancy model (MacKenzie et al., 2002) when the regression effects are modelled using *logistic* or *probit* link functions.

2. Determine the effectiveness of the Variational Bayes algorithms derived under objective 1 in approximating the posterior distributions of the parameters of a nonspatial occupancy model as well as the predictive distribution of the proportion of occupied sites in a region when the regression effects of the model are modelled using *logistic* or *probit* link functions.

3. Develop and formulate a Gibbs sampling algorithm for the single season nonspatial and spatial occupancy models when one uses the *logistic* link function to model the

regression effects of the parameters of both of the models.

4. Evaluate the effectiveness of the Gibbs sampling algorithms developed for the single season nonspatial and spatial occupancy models (SSSO).

5. Develop and formulate Gibbs sampling algorithms for the single season (multi-species) occupancy models when one uses the *logistic* link function to model the regression effects of the parameters of both of a multi-species occupancy model.

1.3 Structure of the thesis

After the introduction to the thesis, the thesis is divided into five chapters. Chapters Two to Five of this thesis are presented as stand-alone academic articles and thus contain some repetition with regards to aspects of the introduction of the chapters. Two chapters (namely Chapters Two and Four) have been published while the remaining will be submitted for peer review at international journals. The candidate is the main author on all articles while supervisor and collaborators appear as co-authors. Each chapter contains an abstract of the chapter at the start and provides an explanation of the contributions made by all authors. To present the thesis as an article-based thesis and to aid readability, all references have been listed alphabetically at the end of the thesis. Most programming codes used during the completion of the thesis have not been included in the manuscript although the reader is directed to its locations where needed.

In Chapters Two and Three I investigate the use of Variational Bayes when applied to a single species (single season) occupancy model (SSO) if the regression effects of the occupancy and detection processes are modelled using *logistic* and *probit* link functions respectively. The algorithms developed are compared to the results obtained when undertaking Markov Monte Carlo methods via the use of simulation studies as well as practical applications of the methods.

Chapter Four commences with a brief discussion of the link between *logistic* regression and occupancy models. This link is exploited to develop a Gibbs sampling algorithm for the SSO model as well as for a particular spatial occupancy model when the regression effects of the occupancy and detection processes are modelled using *logistic* link functions. Via a simulation study, the Gibbs sampling algorithm developed for the SSO model is compared to fitting an SSO model in *JAGS* 4.2.0 and *Stan* 2.17.3, as well as the Gibbs sampling algorithm developed when using the *difference random utility method* developed by Frühwirth-Schnatter and Frühwirth (2007). In this chapter I also analyse two detection-nondetection data sets of South African bird species to illustrate the methods and R package (*Rcppocc)* that was developed to analyse SSO and SSSO models. The package is freely available and can be downloaded from `https://github.com/AllanClark/Rcppocc`.

In Chapter Five I develop Gibbs sampling algorithms that can be used to fit various Bayesian multi-species occupancy models to detection-nondetection data. I illustrate how to fit these models when the detection and occupancy processes are modelled using *logistic* link functions and apply these methods to a camera-trapping study undertaken by Drouilly et al. (2018).

Conclusions and further discussions are undertaken in Chapter six.

Chapter 2

A Variational Bayes approach to the analysis of occupancy models.

The following paper was published in PlosOne in 2016. Minor changes have been made to this version of the published manuscript. They mainly relate to differences between the typographical type-setting used in the published version and this chapter. As two examples, here all vectors and matrices are defined using square braces instead of round braces while L has been changed to $\mathcal{L}$. Here, I also corrected a few grammatical errors that was not picked up during the editorial process. These mainly relate to replacing 'are' with 'is' in certain places.

I have not included all of the R code required to undertake the analysis here. The interested reader can download it from `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148966#sec010` or `https://github.com/AllanClark/PhdAnalysisCode`.

The complete citation of the published manuscript is: Clark A.E., Altwegg R., Ormerod J.T., (2016) A Variational Bayes Approach to the Analysis of Occupancy Models. PLoS ONE 11(2): e0148966. doi:10.1371/journal.pone.0148966.

A portion of the work in this chapter was presented at the *International Statistical Ecology conference*, Montpellier, France, 1-4 July 2014 where I obtained the runner-up position in the PhD. student oral-competition.

Note that the references for the article are displayed at the end of the thesis.

'Allan Ernest Clark' wrote the paper and performed the mathematical development and analysis. 'Res Altwegg', 'John Ormerod' and 'Allan Ernest Clark' reviewed the manuscript extensively. All authors have read and approved the manuscript.

A Variational Bayes approach to the analysis of occupancy models

Allan E. Clark [1,2], Res Altwegg[1,2], John T. Ormerod [3,4]

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.
3. School of Mathematics and Statistics, University of Sydney, Sydney, Australia.
4. ARC Centre of Excellence for Mathematical & Statistical Frontiers, Melbourne, Australia.

Abstract

Detection-nondetection data are often used to investigate species range dynamics using Bayesian occupancy models which rely on the use of Markov chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the parameters of the model. In this article we develop two Variational Bayes (VB) approximations to the posterior distribution of the parameters of a single season site occupancy model which uses logistic link functions to model the probability of species occurrence at sites and of species detection probabilities. This task is accomplished through the development of iterative algorithms that do not use MCMC methods. Simulations and small practical examples demonstrate the effectiveness of the proposed technique. We specifically show that (under certain circumstances) the variational distributions can provide accurate approximations to the true posterior distributions of the parameters of the model when the number of visits per site (K) are as low as three and that the accuracy of the approximations improves as K increases. We also show that the methodology can be used to obtain the posterior distribution of the predictive distribution of the proportion of sites occupied (PAO).

Introduction

Bayesian analysis is a coherent statistical paradigm whereby prior information regarding the research area is blended with that of information obtained from the observed data (Robert and Casella, 1999). Subjective prior information is *elicited* either from expertise in the field or based on prior research (meta analyses). Informative priors are increasingly being used in ecology (Kuhnert et al., 2010; Choy et al., 2009) and even in the absence of prior information many ecologists are using Bayesian methods (Clark, 2005).

One class of model that is often analysed in a Bayesian way is the occupancy model (MacKenzie et al., 2002). The single season occupancy model was formulated by using ideas borrowed from closed population mark-recapture models. In this model n sites are visited a number of times (K) in order to estimate the occupancy ($\boldsymbol{\psi}$) and detection probability (denoted throughout as $\boldsymbol{d}$[1] of a species associated with each site. (The term *detection probability* should be read as *conditional detection probability* throughout the text.) These methods are particularly useful when studying the range dynamics of various animal species and have extensively been applied in the ecological literature (refer to Bailey et al. (2014), Bled et al. (2013) and Broms et al. (2014) for some examples). The model has been formulated as a hierarchical Bayesian model which has lead to numerous extensions of the single season occupancy model (Royle and Dorazio, 2008; Johnson et al., 2013; Royle and Kery, 2007).

Many papers have investigated the statistical properties of the estimators of the single season occupancy model. The first of these developed a maximum likelihood formulation of the model and investigated the properties of the estimators for the occupancy and detection probabilities using simulations. They assume that the parameters of the model are constant for all sites although they also consider incorporating covariates in the model. They found that when $d \geq 0.3$, the parameter estimates of the occupancy probability were reasonably unbiased when $K \geq 5$ while when $K = 2$, a detection probability of at least 0.5 is required to provide a reasonable estimate of ψ. They also found that when the true detection probability is low that $\hat{\psi}$ tends to 1 (MacKenzie et al., 2002). Numerous authors have found similar results regarding boundary problems (Welsh et al., 2013; Guillera-Arroita et al., 2014; Hutchinson et al., 2015) although it has been argued that boundary parameter estimates are rare but could be observed in small data sets (Guillera-Arroita et al., 2014).

Moreno and Lele investigated the small sample properties of the maximum likelihood estimators (Moreno and Lele, 2010). They note that 'When detection or occupancy

[1]Traditionally, the detection probability is denoted as p. We do not use the conventional notation here so as to not create unnecessary confusion with $p(\boldsymbol{x}|\boldsymbol{\theta})$ which denotes the likelihood. Here $\boldsymbol{x}$ denotes the data used in an analysis and $\boldsymbol{\theta}$ is the associated parameters of the assumed model.

probability is small or when the number of sites and number of visits per site is small, maximum likelihood estimators (MLE) of site occupancy parameters have large biases, are numerically unstable, and the corresponding confidence intervals have smaller than nominal coverage.' They proposed a penalized maximum likelihood method which performed adequately for small sample sizes. Recently, their study has been extended by considering three different penalized likelihood type models (Hutchinson et al., 2015). They found that the penalized methods performed well and suggested that 'fully Bayesian methods would be competitive'.

Here, we develop Variational Bayes (VB) approximations to the posterior distribution of the parameters of a single season site occupancy model. One big advantage of the methods developed here is the fact that they could be applied to cases where the researcher has informative priors and might not want to rely on the use of the MLE method. In that situation, Markov Chain Monte Carlo (MCMC) methods were so far the only methods available for fitting occupancy models in a Bayesian analysis. However, for big data sets, MCMC methods can be too slow to be useful. Admittedly the potential computational efficiencies accrued from using a VB algorithm compared to the MLE method would possibly only apply when fitting more complicated occupancy type models. We view our contribution as a first step towards developing similar methods for more complicated occupancy models (e.g., the inclusion of site-specific random effects, spatial occupancy models and dynamic occupancy models).

The proportion of occupied sample locations ($\overline{z} = 1/n \sum_i z_i$, where z_i is the occupancy state for site i) is a derived parameter of interest in many studies (MacKenzie, 2006; Royle and Dorazio, 2008). Although frequentist methods can be used to estimate $\overline{z}$, the calculation of a valid confidence interval for $\overline{z}$ is problematic for the frequentist. The same holds true for prediction of occupancy status in species distribution models (Kéry et al., 2010). We show (via simulations as well as using practical examples) that the VB approximations can be used to accurately obtain prediction intervals for latent state variables (e.g., occupancy states $\boldsymbol{z}$) or for functions of these state variables by simulating from the VB posterior distributions.

This paper commences with a brief discussion of Variational Bayes (VB). Thereafter, a VB implementation of a particular occupancy model is developed in Section 2.1.2. In Section 2.1.3 the results of a short simulation study are presented while in Section 2.2.2 we analyse site occupancy data of five bird species to illustrate the usefulness of the VB technique developed. A list of some of the notations and distribution theory used in the text can be found in the Supporting information[2] on page 32.

[2]The phrase 'Supporting information' can be read as 'Appendix' and is used here due to the conventions of the journal as at publication date.

2.1 Material and Methods

2.1.1 A brief introduction to Variational Bayes (VB)

Variational Bayes is used to approximate posterior distributions obtained when undertaking Bayesian analysis and could be useful in many ecological applications.

In what follows let $\boldsymbol{\theta}$ be a vector of parameters of a statistical model, $\pi(\boldsymbol{\theta})$ be a prior distribution for these parameters and $\boldsymbol{y}$ be a random variable. In the context of this article, $\boldsymbol{\theta}$ are the parameters of a single season occupancy model while $\boldsymbol{y}$ represents detection-nondetection data used to fit an occupancy model. Further, suppose that a posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is not analytically tractable and that analytical expressions for its posterior moments do not exist. In probability theory the Kullback-Leibler (KL) divergence provides a measure of a difference between two probability distributions (Kullback and Leibler, 1951). When the two distributions being compared are exactly the same the divergence measure is equal to zero while when they are different the divergence measure is positive.

The VB method approximates a posterior distribution by using a distribution $q(\boldsymbol{\theta})$ which is obtained by minimizing the Kullback-Leibler (KL) divergence between $q(\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. The KL divergence is

$$KL(q(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}|\boldsymbol{y})) = \int q(\boldsymbol{\theta}) \ln\left(\frac{q(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\boldsymbol{y})}\right) d\boldsymbol{\theta} = \int q(\boldsymbol{\theta}) \ln\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta},\boldsymbol{y})}\right) d\boldsymbol{\theta} + \ln p(\boldsymbol{y}) \quad (2.1)$$

where $p(\boldsymbol{y})$ is the marginal likelihood, $p(\boldsymbol{\theta},\boldsymbol{y})$ is the joint likelihood of the data and the parameter vector $\boldsymbol{\theta}$ with

$$\mathcal{L}(q(\boldsymbol{\theta})) = \int q(\boldsymbol{\theta}) \ln\left(\frac{p(\boldsymbol{\theta},\boldsymbol{y})}{q(\boldsymbol{\theta})}\right) d\boldsymbol{\theta}. \quad (2.2)$$

Since $KL(q(\boldsymbol{\theta})||\pi(\boldsymbol{\theta}|\boldsymbol{y})) \geq 0$, $\ln p(\boldsymbol{y}) \geq \mathcal{L}(q(\boldsymbol{\theta}))$ for every $q(\boldsymbol{\theta})$ and minimising $KL(q(\boldsymbol{\theta})||$ $p(\boldsymbol{\theta}|\boldsymbol{y}))$ is equivalent to maximising $\mathcal{L}(q(\boldsymbol{\theta}))$. Often it is assumed that $q(\boldsymbol{\theta})$ can be factorized as a product of simple probability distributions as $q(\boldsymbol{\theta}) = \prod_i q(\theta_i)$ where each of the $q(\theta_i)$ are iteratively estimated as $\ln q(\theta_i) \propto \mathbb{E}_{-\theta_i}(\ln p(\boldsymbol{\theta},\boldsymbol{y}))$. Here $\mathbb{E}_{-\theta_i}$ denotes an expectation with respect to the density $\prod_{j\neq i} q(\theta_j)$. An alternate method of obtaining $q(\boldsymbol{\theta})$ involves making an assumption regarding its parametric form. The parameters of this distribution are obtained by maximising $\mathcal{L}(q(\boldsymbol{\theta}))$ (Ormerod and Wand, 2010).

VB is often used as an alternative to MCMC methods since the method can be much faster to implement since in most applications $q(\theta_i)$ will be of a known simple form (McGrory and Titterington, 2007; Bishop, 2008; Hensman et al., 2012). Variational approximations to posterior distributions can accurately estimate the posterior mean of the parameters,

although the posterior variances of some of the parameters might be underestimated (Wang and Titterington, 2005; Grimmer, 2010). Although this problem is context specific the estimate of the posterior variance is asymptotically valid for linear models (You et al., 2014). As a solution the variational covariance matrix is often replaced by the inverse of the Fishers' information matrix (Wang and Titterington, 2005). Alternatively the non-parametric bootstrap could be used to provide interval estimates of the parameters (Nathoo et al., 2014).

2.1.2 VB applied to single season occupancy models

In a single season occupancy model n sites are visited $\boldsymbol{K}$ times in order to estimate the occupancy ($\boldsymbol{\psi}$) and detection ($\boldsymbol{d}$) probability of a species associated with each site. Each site could be surveyed a different amount of times such that $\boldsymbol{K} = [K_1, K_2, \ldots, K_n]^T$ where K_i represents the number of surveys undertaken to site i and $\boldsymbol{d}$ is a ragged matrix with dimensions determined by $\boldsymbol{K}$. The total number of site visits undertaken is defined as $N = \sum_{i=1}^{n} K_i$.

The data collected at each site are represented as an N dimensional vector $\boldsymbol{y} = [\boldsymbol{y}_1^T, \ldots, \boldsymbol{y}_n^T]^T$, where each of the $\boldsymbol{y}_i$ denotes the vector of detections and nondetections for site i. A 0 in the vector $\boldsymbol{y}_i$ indicates that the species was not observed at the i^{th} site during a particular visit while a 1 indicates that the species was observed at the particular site during a particular visit. Let the vector $\boldsymbol{z}$ represent the true species occupancy at the sites considered. Since we are using a single season model, $\boldsymbol{z}$ is assumed to be constant across the season. $\boldsymbol{z}$ is partially observed, i.e. $z_i = 1$ if the species occupies site i and $z_i = 0$ if it does not occupy site i. We know $z_i = 1$ if the species is observed at site i during any of the visits since we assume that there are no false identifications of individuals. If the species is however not observed at site i, z_i could equal 0 or 1 since we are uncertain about whether the species actually occurs at that site. We treat $\boldsymbol{y}_i$ as a row vector and is of dimension $1 \times K_i$ while $\boldsymbol{z}$ is of dimension $n \times 1$.

The single season occupancy models can be represented using the following hierarchical model (Royle and Dorazio, 2008)

$$z_i|\psi_i \sim \text{Bernoulli}(\psi_i)$$
$$y_{i,j}|z_i, d_{i,j} \sim \text{Bernoulli}(z_i d_{i,j})$$

for all sites $i = 1, \ldots, n$; for all visits $j = 1, \ldots, K_i$. $\psi_i = \Pr(z_i = 1)$ denotes the probability that the species occurs at site i while $d_{i,j} = \Pr(y_{i,j} = 1|z_i = 1)$ denotes the conditional probability of detecting the species during the j^{th} visit of site i given that the species is present at site i. The occupancy probabilities and the detection probabilities can be estimated using either maximum likelihood (MacKenzie et al., 2002), penalized maximum

likelihood (Moreno and Lele, 2010) or Bayesian methods (Royle and Dorazio, 2008). In what follows we develop a VB approach to estimating these quantities.

Additional covariate data collected at each of the sites are used to estimate the site occupancy and detection probabilities. Specifically we assume that we have r occupancy and s detection covariates. We further assume that we have no missing values in these covariates. Formally we let $\boldsymbol{W}$ and $\boldsymbol{X}$ be the design matrices for the detection and occupancy effects respectively, with dimensions $N \times s$ and $n \times r$. Correspondingly, let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ be the detection and occupancy effects with dimensions $s \times 1$ and $r \times 1$ respectively. The matrix $\boldsymbol{W}$ is constructed by row-binding the detection covariates at the different locations and for different visits one below each other such that $\boldsymbol{W} = [\underline{\boldsymbol{w}}_1^T, \ldots, \underline{\boldsymbol{w}}_n^T]^T$ where each of the $\underline{\boldsymbol{w}}_i$ matrices are of dimension $K_i \times s$ with $\underline{\boldsymbol{w}}_i = [\boldsymbol{w}_{i,1}^T, \boldsymbol{w}_{i,2}^T, \ldots, \boldsymbol{w}_{i,K_i}^T]^T$.

The occupancy and detection probabilities at the various sites for all visits are modelled using the following logistic link functions

$$\psi_i = \left(1 + \exp\left(-\boldsymbol{x}_i^T \boldsymbol{\beta}\right)\right)^{-1}$$
$$d_{i,j} = \left(1 + \exp\left(-\boldsymbol{w}_{i,j}^T \boldsymbol{\alpha}\right)\right)^{-1}.$$

It can be shown that the conditional likelihood of the data and the true occupancy variables is

$$p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\alpha}, \boldsymbol{\beta}) = \prod_{i=1}^{n} \psi_i^{z_i} \left(1 - \psi_i\right)^{1-z_i} \prod_{j=1}^{K_i} \left(z_i d_{i,j}\right)^{y_{i,j}} \left(1 - z_i d_{i,j}\right)^{1-y_{i,j}}.$$

We now assume that the prior distribution for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are multivariate Gaussian distributions (denoted as $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta})$) with parameters $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}^0$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0$ respectively. We further assume that the **variational approximate distribution** of $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z})$ is of the form $q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}) = q(\boldsymbol{\alpha}, \boldsymbol{\beta}) \prod_i q(z_i)$ where each of the $q(z_i)$ are Bernoulli distributed with success probability $(sp)_i$. Under this restriction $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be factorized into two separate factors $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ with

$$q(\boldsymbol{\alpha}) \propto \exp\left(\boldsymbol{y}^T \tilde{\boldsymbol{P}} \boldsymbol{W} \boldsymbol{\alpha} - \tilde{\boldsymbol{p}}^T b\left(\boldsymbol{W}\boldsymbol{\alpha}\right) + \ln \pi(\boldsymbol{\alpha})\right), \tag{2.3}$$
$$q(\boldsymbol{\beta}) \propto \exp\left(\boldsymbol{p}^T \boldsymbol{X} \boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\beta})\right), \tag{2.4}$$

where $\tilde{\boldsymbol{p}} = \mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{\beta})}(\tilde{\boldsymbol{z}}|\boldsymbol{y})$, $\tilde{\boldsymbol{P}} = \text{diag}(\tilde{\boldsymbol{p}})$, $\boldsymbol{p} = \mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{\beta})}(\boldsymbol{z}|\boldsymbol{y})$ and $b(x) = \ln(1 + \exp(x))$. Here $\tilde{\boldsymbol{z}} = [z_1 \mathbf{1}_{K_1}^T, \ldots, z_n \mathbf{1}_{K_n}^T]^T$. Refer to Section 2.B in the Supplementary information for a derivation of the above results.

The normalization constant of $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is not known analytically and thus $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ are not of a known type. We attempt to approximate the posterior distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using two different methods. In the first method we approximate the variational distribution by using a Laplace approximation to equations (2.3) and (2.4) and thus assume that the

variational distributions are multivariate Gaussian with parameters $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ respectively; while in the second method we employ a tangent based approximation to $b(\boldsymbol{W}\boldsymbol{\alpha})$ and $b(\boldsymbol{X}\boldsymbol{\beta})$ to obtain approximations to $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ respectively.

Once we have obtained approximations to $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ it then follows that the q-densities, $q(z_i|\boldsymbol{y}_i = \mathbf{0})\ \forall i$, are Bernoulli distributed with success probability $(1 + \exp(-c_i))^{-1}$. The approximate conditional occupancy probabilities for all sites can then be calculated for the two methods (denoted as 'L' and 'T' respectively) using

$$c_i^{(L)} = \boldsymbol{x}_i^T \boldsymbol{\mu}_{\boldsymbol{\beta}} - \mathbf{1}_{K_i}^T \mathbb{E}_{q(\boldsymbol{\alpha})} \left(b\left(\underline{\boldsymbol{w}}_i \boldsymbol{\alpha}\right)\right), \tag{2.5}$$

$$\boldsymbol{d}_i = \underline{\boldsymbol{w}}_i^T \mathrm{diag}(A(\underline{\boldsymbol{a}}_i)) \underline{\boldsymbol{w}}_i \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} + \boldsymbol{\mu}_{\boldsymbol{\alpha}} \boldsymbol{\mu}_{\boldsymbol{\alpha}}^T \right) \text{ and}$$

$$c_i^{(T)} = \boldsymbol{x}_i^T \boldsymbol{\mu}_{\boldsymbol{\beta}} + \mathbf{1}_{K_i}^T C(\underline{\boldsymbol{a}}_i) - \frac{1}{2} \mathbf{1}_{K_i}^T \underline{\boldsymbol{w}}_i \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \mathrm{tr}\left(\boldsymbol{d}_i\right). \tag{2.6}$$

Here $b\left(\underline{\boldsymbol{w}}_i \boldsymbol{\alpha}\right)$ is a vector of length K_i such that

$$\left\langle b\left(\underline{\boldsymbol{w}}_i \boldsymbol{\alpha}\right)\right\rangle = \left[\left\langle b\left(\boldsymbol{w}_{i,1}^T \boldsymbol{\alpha}\right)\right\rangle, \left\langle b\left(\boldsymbol{w}_{i,2}^T \boldsymbol{\alpha}\right)\right\rangle, \ldots, \left\langle b\left(\boldsymbol{w}_{i,K_i}^T \boldsymbol{\alpha}\right)\right\rangle \right]^T$$

The parameters of the variational distributions are all dependent on one another and can be computed using an iterative scheme such as that given in Algorithm 1 and Algorithm 2. A detailed description of aspects of the above derivations can be found in the supplemental information to this chapter. In particular, the quantities used to calculate $c_i^{(T)}$ can be found in the Supporting information (on page 33) while an explanation regarding the stopping rule for both algorithms is described on page 36).

Algorithm 1. Iterative scheme for obtaining the parameters of the optimal density of $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using the Laplace approximation.

1. Initialize $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$
2. Cycle:
   2.1 Cycle:
   $$\boldsymbol{g}_1 \leftarrow \boldsymbol{W}^T \left(\tilde{\boldsymbol{P}}\boldsymbol{y} - \tilde{\boldsymbol{p}} \odot b'\left(\boldsymbol{W}\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right) - \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0\right)^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}^0\right)$$
   $$\boldsymbol{g}_2 \leftarrow \boldsymbol{X}^T \left(\boldsymbol{p} - b'\left(\boldsymbol{X}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right) - \left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0\right)^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{\beta}} - \boldsymbol{\mu}_{\boldsymbol{\beta}}^0\right)$$
   $$\boldsymbol{\Sigma}_{11} \leftarrow \left(\boldsymbol{W}^T \mathrm{diag}\left(\tilde{\boldsymbol{p}} \odot b''\left(\boldsymbol{W}\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right) \boldsymbol{W} + \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0\right)^{-1} \right)^{-1}$$
   $$\boldsymbol{\Sigma}_{22} \leftarrow \left(\boldsymbol{X}^T \mathrm{diag}\left(b''\left(\boldsymbol{X}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right) \boldsymbol{X} + \left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0\right)^{-1} \right)^{-1}$$
   $$\boldsymbol{\mu}_{\boldsymbol{\alpha}} \leftarrow \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Sigma}_{11}\boldsymbol{g}_1 \text{ and } \boldsymbol{\mu}_{\boldsymbol{\beta}} \leftarrow \boldsymbol{\mu}_{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{22}\boldsymbol{g}_2$$
   until the Newton-Raphson algorithm converges.
   2.2 Calculate conditional occupancy probabilities for all sites where $\boldsymbol{y}_i = \mathbf{0}$ using equation (2.5). Note that $(sp)_i = 1$ for all sites where $\underline{\boldsymbol{y}}_i \neq \underline{\mathbf{0}}$.

until the change in $\mathbb{E}_q(\ln \underline{p}) - \mathbb{E}_q(\ln q(\boldsymbol{\alpha}, \boldsymbol{\beta})) - \mathbb{E}_q(\ln q(\boldsymbol{z}))$ becomes negligible. ($\leq 10^{-6}$)

Algorithm 2. Iterative scheme for obtaining the parameters of the optimal density of $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using the tangent based method.

1. Initialize $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$, $\boldsymbol{a}_N > 0$ and $\boldsymbol{b}_N > 0$.
2. Cycle:
   2.1 Calculate the conditional occupancy probabilities for all sites where $\boldsymbol{y}_i = \boldsymbol{0}$ using equation (2.6). Note that $(sp)_i = 1$ for all sites where $\underline{\boldsymbol{y}}_i \neq \underline{\boldsymbol{0}}$.
   2.2 Set $\boldsymbol{\mu}_{\boldsymbol{\alpha}} \leftarrow \boldsymbol{B}_1^{-1}\boldsymbol{B}_2^T$, $\boldsymbol{\mu}_{\boldsymbol{\beta}} \leftarrow \boldsymbol{D}_1^{-1}\boldsymbol{D}_2^T$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \leftarrow \boldsymbol{B}_1^{-1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} \leftarrow \boldsymbol{D}_1^{-1}$ with
   $$\boldsymbol{B}_1 \leftarrow (\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1} - 2\boldsymbol{W}^T \text{diag}(A(\boldsymbol{a}) \odot \tilde{\boldsymbol{p}})\boldsymbol{W}$$
   $$\boldsymbol{B}_2 \leftarrow \left(\tilde{\boldsymbol{P}}\boldsymbol{y} - \tfrac{1}{2}\tilde{\boldsymbol{p}}\right)^T \boldsymbol{W} + (\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0)^T(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1}$$
   $$\boldsymbol{D}_1 \leftarrow (\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1} - 2\boldsymbol{X}^T \text{diag}(A(\boldsymbol{b}))\boldsymbol{X}$$
   $$\boldsymbol{D}_2 \leftarrow \left(\boldsymbol{p} - \tfrac{1}{2}\boldsymbol{1}_n\right)^T \boldsymbol{X} + (\boldsymbol{\mu}_{\boldsymbol{\beta}}^0)^T(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1}$$
   2.3 Calculate the 'variational parameters'. Refer to page 33.

until the change in $\mathbb{E}_q(\ln \underline{p}) - \mathbb{E}_q(\ln q(\boldsymbol{\alpha}, \boldsymbol{\beta})) - \mathbb{E}_q(\ln q(\boldsymbol{z}))$ becomes negligible. ($\leq 10^{-6}$)

2.1.3 Simulation Study

In the following simulation study we investigate some of the properties of the VB method and investigate whether it could be used to produce *statistically valid inference*. We specifically focus on the frequentist properties of the posterior mean parameters of the VB distribution of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. This task is undertaken by *empirically* comparing the coverage probability and credible/confidence intervals of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ associated with the two VB methods developed and comparing these to the same statistics obtained using MCMC and maximum likelihood. We calculate credible intervals for the Bayesian methods and confidence intervals for the MLE method and focus particularly on the 95% credible or confidence intervals.

The accuracy of the VB approximations to the posterior distribution obtained through MCMC is also assessed. This is undertaken by calculating $\text{acc}(x) = 1 - \frac{1}{2}\int |q(x) - q_{MCMC}(x)|dx$. The $acc(x)$ measure lies between 0 and 1 with a value of 1 indicating a perfect approximation and a value close to 0 indicating a poor approximation by the variational distribution to the true posterior distribution.

Occupancy models are often used to assess the predictive distribution of the proportion of occupied sites defined as $\text{PAO} = \frac{1}{n}\sum_{i=1}^n z_i$. We thus investigate the posterior approximation of the PAO using the Laplace VB posterior approximation method. These can easily be obtained by sampling from the VB posterior distribution for each z_i in turn to construct the PAO statistic. To assess the VB approximations the $\text{acc}(x)$ statistic was used.

We consider 32 simulation settings. The number of sites (n) is set to 50 and 100 while the number of visits to each site (K) is set to $2, 3, 4$ and 5 respectively. The following combinations of the regression coefficients were used: 1. $\boldsymbol{\alpha} = [0,\ 1.75]^T$, $\boldsymbol{\beta} = [-1.85,\ 2.5]^T$; 2. $\boldsymbol{\alpha} = [1.35,\ 1.75]^T$, $\boldsymbol{\beta} = [-1.85,\ 2.5]^T$; 3. $\boldsymbol{\alpha} = [0,\ 1.75]^T$ $\boldsymbol{\beta} = [-0.1,\ 2.5]^T$ and 4.

$\boldsymbol{\alpha} = [1.35,\ 1.75]^T$ $\boldsymbol{\beta} = [-0.1,\ 2.5]^T$. These parameter values ensure an approximate average detection and occupancy probability among the sites of (0.5, 0.3), (0.7, 0.3), (0.5, 0.5) and (0.7, 0.5) respectively. We have not considered any cases where the detection and occupancy probabilities are lower than 0.3 since in these cases data sets are expected to be very sparse which requires many site visits in order to undertake useful statistical inference (MacKenzie et al., 2002).

The occupancy regression covariate was obtained by standardising a Uniform$(-2, 2)$ random variable while the detection covariate was obtained by standardising a Uniform$(-5, 5)$ random variable. Each of these variables were transformed to have a zero mean and a standard deviation of one. The following parameter vectors were used to specify the prior distribution of the parameters: $\boldsymbol{\mu}_i^0 = [0, 0]^T$, $\boldsymbol{\Sigma}_i^0 = \text{diag}[1000, 1000]$ for $i = \boldsymbol{\alpha}, \boldsymbol{\beta}$.

Each simulation setting was replicated 350 times. All calculations were undertaken using R 3.3.1 (R Core Team, 2014). Numerical optimizations were performed using the the BFGS method of the R function *optim*; MCMC sampling was undertaken using the R package R2jags (Su and Yajima, 2012) in combination with JAGS 3.4.0 (Plummer, 2003) while all variational approximations were performed using the authors' code. 100 000 posterior samples were obtained for each MCMC simulation. The first 25 000 samples were discarded as burn-in samples while the remaining 75 000 samples were retained. Prior experimentation using the MCMC algorithm indicated that 25 000 iterations are enough to ensure that the Markov chains would converge to the stationary distributions. The posterior samples were not thinned (Link and Eaton, 2012).

2.2 Results

2.2.1 Simulation Results

Table 1 contains a summary of some of the results of the simulation study. For each value of K we tabulate the median coverage probability and credible/confidence interval width of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ associated with the four[3] estimation procedures considered here. The medians are calculated across the different occupancy and detection probability combinations for fixed values of n and K.

We found that as the number of sites increased, the credible (and confidence) interval widths of the true regression parameters decreased (for all methods). For a fixed number of sites, the credible (and confidence) interval widths of the true $\boldsymbol{\alpha}$ parameter values decreased as K increased while the associated widths for the $\boldsymbol{\beta}$ parameters did not appear

[3]The MCMC estimation results were removed from Table 1 since the coverage is always 95%.

Table 1. The median coverage probability and credible/confidence interval widths of the covariate effects for the single season occupancy model.

			K							
			2		**3**		**4**		**5**	
n	Param	Method	Coverage	Width	Coverage	Width	Coverage	Width	Coverage	Width
50	α_0	Laplace	0.929	2.006	**0.944**	1.619	**0.956**	1.393	**0.962**	1.221
		Tangent	0.793	**1.51**	0.838	**1.22**	0.841	**1.05**	0.854	**0.93**
		MLE	**0.96**	2.675	0.964	1.869	0.962	1.509	0.968	1.28
	α_1	Laplace	**0.944**	2.452	**0.959**	1.986	**0.96**	1.704	0.966	1.493
		Tangent	0.76	**1.642**	0.787	**1.31**	0.812	**1.127**	0.803	**0.996**
		MLE	0.962	2.926	0.97	2.215	0.968	1.81	**0.963**	1.55
	β_0	Laplace	**0.936**	1.994	**0.952**	2.019	**0.963**	2.032	**0.964**	2.022
		Tangent	0.75	**1.345**	0.821	**1.348**	0.845	**1.349**	0.841	**1.349**
		MLE	0.973	2.776	0.974	2.505	0.974	2.416	0.97	2.322
	β_1	Laplace	0.92	2.656	**0.952**	2.679	**0.952**	2.705	**0.952**	2.71
		Tangent	0.666	**1.493**	0.744	**1.494**	0.733	**1.498**	0.723	**1.498**
		MLE	**0.954**	4.11	0.963	3.578	0.964	3.419	0.969	3.28
100	α_0	Laplace	0.907	1.402	0.932	1.103	**0.946**	0.944	0.942	0.852
		Tangent	0.764	**1.048**	0.823	**0.842**	0.858	**0.726**	0.826	**0.651**
		MLE	**0.96**	1.784	**0.96**	1.24	**0.954**	1.008	**0.95**	0.88
	α_1	Laplace	**0.949**	1.728	0.942	1.351	**0.949**	1.143	**0.953**	1.033
		Tangent	0.76	**1.126**	0.768	**0.905**	0.806	**0.778**	0.79	**0.696**
		MLE	0.968	2.025	**0.949**	1.444	0.947	1.182	**0.953**	1.05
	β_0	Laplace	0.928	1.459	**0.957**	1.478	**0.95**	1.475	**0.95**	1.484
		Tangent	0.723	**0.954**	0.766	**0.957**	0.792	**0.956**	0.791	**0.958**
		MLE	**0.96**	1.931	0.968	1.729	0.968	1.638	0.96	1.607
	β_1	Laplace	0.919	1.945	**0.944**	1.979	**0.947**	1.969	**0.95**	2.01
		Tangent	0.623	**1.052**	0.654	**1.058**	0.71	**1.055**	0.67	**1.06**
		MLE	**0.96**	2.828	0.963	2.482	0.958	2.292	0.954	2.225

The medians are calculated across the different occupancy and detection probability combinations for fixed values of n and K. n represents the total number of sites visited while K represents the number of visits to each site. We highlight the method with the smallest credible/confidence interval width as well the method (not considering the MCMC method) with the closest coverage probability to 0.95.

to decrease noticeably with an increase in K. The simulation results suggests that the coverage probabilities associated with the Laplace method and the MLE methods (for all regression parameters) are very close to that of the nominal coverage value of 0.95 for $K \geq 3$. It is evident from these results that the Tangent based method does not perform well under any of the scenarios considered and consistently produced the smallest credible interval widths.

Based on the accuracy calculations across the replicate data sets, the Tangent based method generally appears to be worse at approximating the marginal posterior distributions of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ when comparisons are made based on the median accuracy measure for

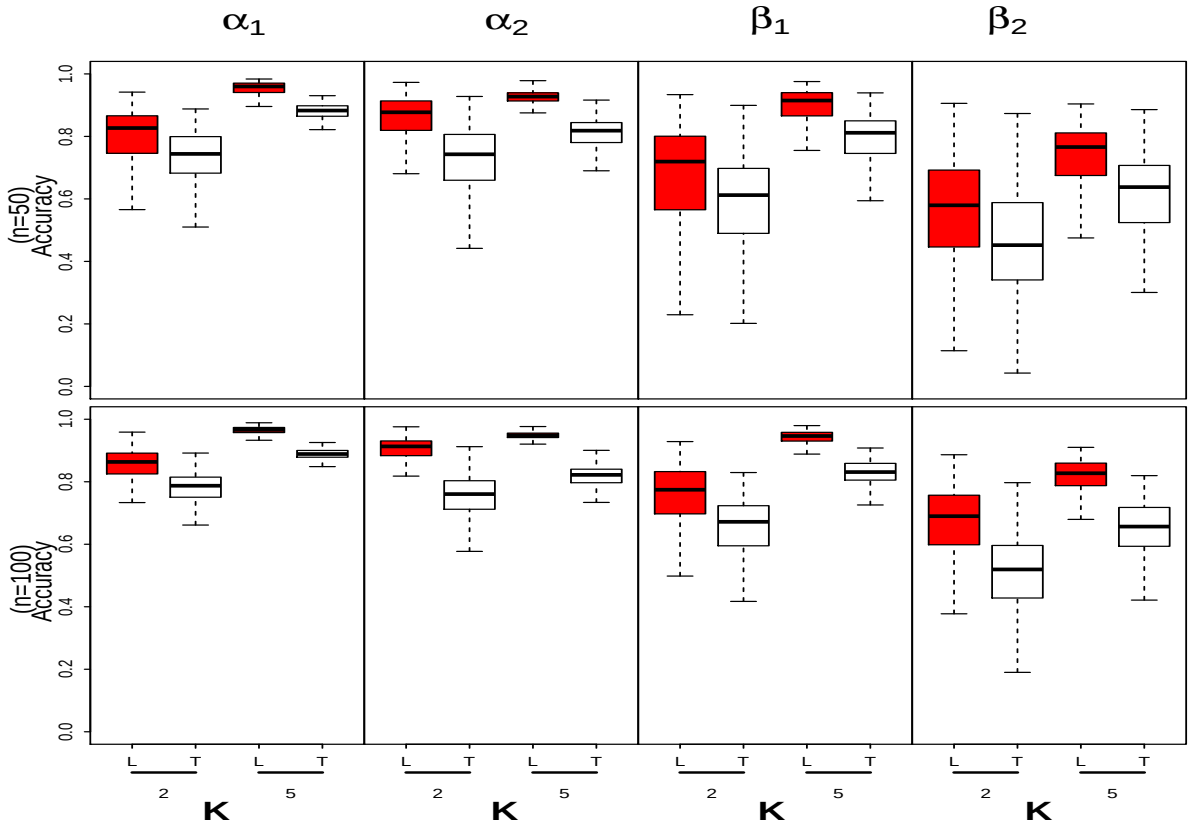

Figure 3. Box plots of the accuracy measurements for the model parameters associated with the Laplace (L-dark boxes) and Tangent (T-light boxes) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. The detection and occupancy probabilities are approximately 0.5. The accuracy of the VB approximations is measured by calculating $\text{acc}(x) = 1 - \frac{1}{2}\int |q(x) - q_{MCMC}(x)|dx$. The measure lies between 0 and 1 with a value of 1 indicating a perfect approximation and a value close to 0 indicating a poor approximation by the variational distribution to the true posterior distribution.

these parameters. As an example of these simulation results, consider the scenario where the estimated mean detection and occupancy probability across all sites and revisits are both 0.5 (refer to Figure 3). In general, the posterior approximations for the detection regression parameters were quite good with median accuracy statistics greater than 0.8 even when the number of revisits are small. The accuracy statistics for the detection regression parameters dramatically increase as n and K increases with median accuracy statistics in excess of 0.95 when $K = 5$. Similar comments can be made regarding the accuracy of the posterior approximations for the occupancy regression parameters. In general, the accuracies increase with an increase in n and K however the rate of increase in the accuracy statistics for the occupancy regression parameters appears slower than those observed for the detection regression parameters. The box plots of the accuracy statistics associated with the different methods for the remaining cases can be found in the Supplementary information (refer to Figures 7, 8 and 9).

We found that the accuracy of the approximate predictive distributions for the proportion of occupied sites improves as K increases (refer to Figure 4). This observation is consistent across all of the scenarios considered. From an examination of the posterior predictive distributions (not shown here) it is evident that the VB predictive distributions are lighter tailed than the MCMC predictive distributions however this effect is reduced for $K \geq 3$. This observation can clearly be seen when examining the results displayed in Table 2 and Table 3. It is noticeable that the summary statistics of the predictive distributions using the two methods are very similar although the VB predictive distributions display a slightly reduced posterior variance under certain conditions.

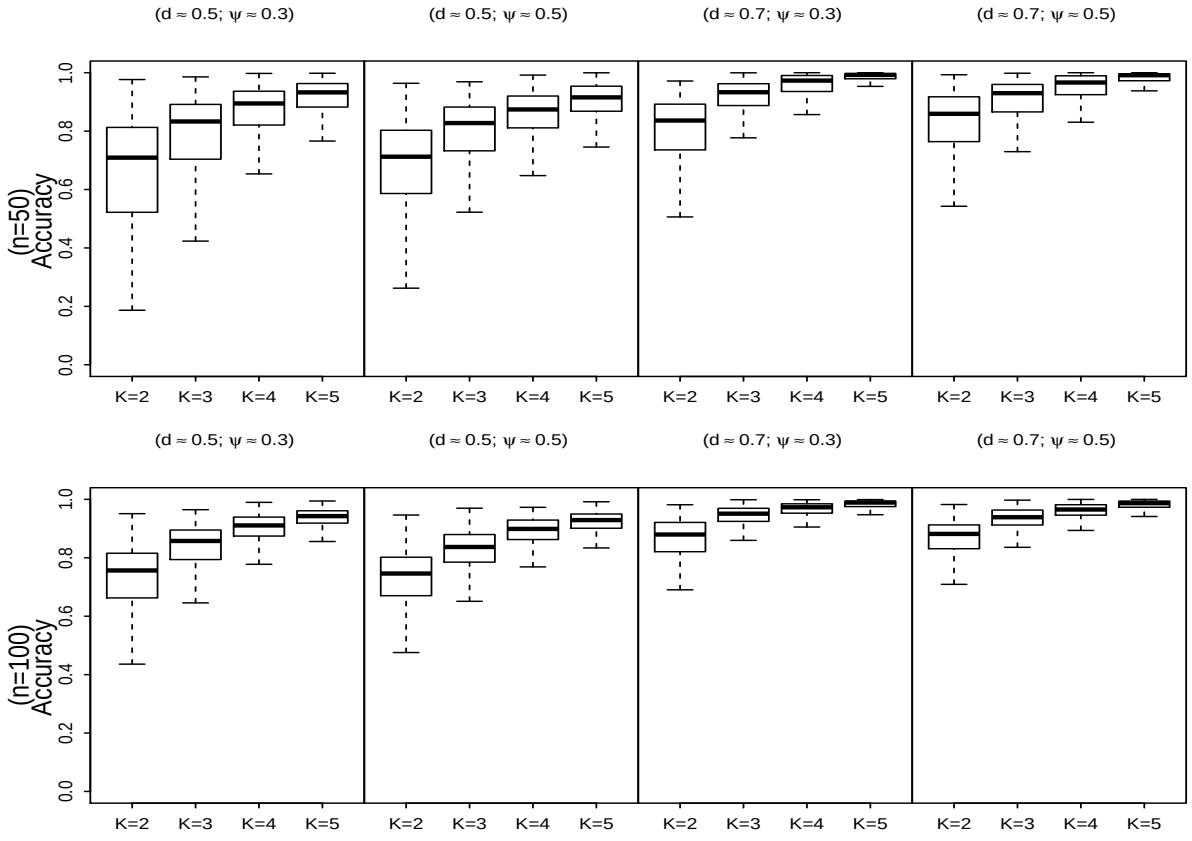

Figure 4. Box plots of the accuracy measurements for the predictive distribution of the proportions of occupied sites associated with the Laplace method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 3,\ 4$ and 5. The accuracy of the VB approximations is measured by calculating $\text{acc}(x) = 1 - \frac{1}{2}\int |q(x) - q_{MCMC}(x)|dx$. The measure lies between 0 and 1 with a value of 1 indicating a perfect approximation and a value close to 0 indicating a poor approximation by the variational distribution to the true posterior distribution.

Table 2. Summary statistics of the posterior predictive distributions of the proportion of sites occupied using the MCMC method and the VB method for different scenarios. ($n = 50$)

		MCMC Method					**Laplace Method**				
Case	K	Mean	Std	Median	2.5%	97.5%	Mean	Std	Median	2.5%	97.5%
	2	16.8	5.2	16	8	29	14.6	3.3	15	8	21
$d \approx 0.5$	3	15.6	3.6	15	10	24	14.6	2.9	14	9	20
$\psi \approx 0.3$	4	15.0	2.8	15	10	21	14.6	2.6	15	10	20
	5	14.9	2.6	15	10	20	14.7	2.5	15	10	20
	2	25.7	4.7	25	18	36	23.9	3.5	24	17	31
$d \approx 0.7$	3	25.0	3.6	25	18	33	24.1	3.1	24	18	30
$\psi \approx 0.3$	4	25.0	3.0	25	19	31	24.6	2.8	25	19	30
	5	24.7	2.8	25	20	30	24.5	2.7	24	19	30
	2	15.8	3.5	15	10	24	14.9	2.7	15	10	20
$d \approx 0.5$	3	14.8	2.5	15	10	20	14.5	2.4	14	10	20
$\psi \approx 0.5$	4	14.5	2.4	14	10	19	14.4	2.4	14	10	19
	5	14.6	2.4	15	10	19	14.6	2.4	15	10	19
	2	25.2	3.3	25	19	32	24.6	2.9	25	19	30
$d \approx 0.7$	3	24.6	2.8	25	20	30	24.4	2.7	24	19	30
$\psi \approx 0.5$	4	24.6	2.5	25	20	30	24.6	2.5	25	20	29
	5	24.7	2.5	25	20	29	24.6	2.5	25	20	29

Table 3. Summary statistics of the posterior predictive distributions of the proportion of sites occupied using the MCMC method and the VB method for different scenarios. ($n = 100$)

		MCMC Method					**Laplace Method**				
Case	K	Mean	Std	Median	2.5%	97.5%	Mean	Std	Median	2.5%	97.5%
	2	31.0	6.3	30	21	45	29.4	4.6	29	21	39
$d \approx 0.5$	3	30.4	4.6	30	22	40	29.7	4.1	30	22	37
$\psi \approx 0.3$	4	29.7	3.8	30	22	37	29.4	3.6	29	22	36
	5	29.7	3.6	30	22	37	29.5	3.5	29	22	36
	2	49.7	6.3	49	38	63	48.3	5.2	48	38	58
$d \approx 0.7$	3	49.6	4.9	50	40	59	48.9	4.6	49	40	58
$\psi \approx 0.3$	4	49.3	4.2	49	41	58	48.9	4.1	49	41	57
	5	49.1	3.9	49	41	57	48.9	3.9	49	41	56
	2	30.2	4.4	30	22	40	29.7	4.0	30	22	38
$d \approx 0.5$	3	29.6	3.5	30	23	36	29.5	3.4	30	23	36
$\psi \approx 0.5$	4	29.5	3.3	30	23	36	29.4	3.3	29	23	36
	5	29.5	3.3	29	23	36	29.5	3.2	29	23	36
	2	49.4	4.5	49	41	58	48.9	4.2	49	41	57
$d \approx 0.7$	3	49.2	3.8	49	42	57	49.0	3.8	49	42	56
$\psi \approx 0.5$	4	49.1	3.7	49	42	56	49.0	3.7	49	42	56
	5	49.0	3.7	49	41	56	49.0	3.7	49	41	56

2.2.2 Application to real data sets

As examples of the proposed technique, we use detection-nondetection data extracted from the second Southern African Bird Atlas Project (Harebottle et al., 2007) database

(refer to `http://sabap2.adu.org.za/`) for 2012 to compare the performance of different methods for fitting a single season occupancy model. The data were collected by citizen scientists using 5-minute latitude $\times$ 5-minute longitude rectangular grids across South Africa (Harebottle et al., 2007). Each site is approximately 8 km $\times$ 7.6 km (Broms et al., 2014). The citizen scientists were asked to make a list of all the species that they encountered during at least two hours of intense birding. They were allowed to add additional species to the list for up to five days. By providing information on the species that they encountered, the citizen scientists implicitly also provided information about the species they did not encounter. Hence, we extracted detection-nondetection data for five bird species (1. Black-headed heron (*Ardea melanocephala*), 2. Egyptian goose (*Alopochen aegyptiaca*), 3. orange-throated longclaw (*Macronyx capensis*), 4. white-browed sparrow-weaver (*Plocepasser mahali*) and 5. Long-tailed widowbird (*Euplectes progne*)) from this database, treating each check-list as an independent observation. We included all grid cells in and around Gauteng, South Africa, that contained a minimum of three site visits. Many of the sites were visited a large number of times but we limited the maximum number of site visits to five (since the focus of the analysis was to assess whether the VB techniques could be used to analyse studies which have relatively small sample sizes and low number of revisits per site). This restriction reduced the data sets to 123 sites; 50 of which had three surveys; 52 had four surveys and the remaining 21 sites had 5 surveys.

In our analysis we specifically compare the MLE, MCMC and the VB methods where uninformative priors (as in the simulation study) were used for all parameters. We fitted a model with one detection covariate and one occupancy covariate. The detection covariate used was the number of species observed by the birder (denoted as *nspp*) while the occupancy probability was modelled as a function of the ratio of potential to realised evapotranspiration (*AETdivPETs*). *AETdivPETs* is a measure of vegetation cover and hydric stress and is an important predictor for bird species occurrence in South Africa (Péron and Altwegg, 2015). Both covariates were standardized to have zero mean and unit variance.

Maximum likelihood estimation was undertaken using the R package *unmarked* (Fiske and Chandler, 2011); MCMC sampling was undertaken using the R package jagsUI (Kellner, 2014) while all variational approximations were performed using the authors' code. The R code used to perform the analysis (refer to page 40), the data as well as documentation regarding the VB code (refer to pages 40 and 48 respectively) can be found in the Supporting information. The MCMC estimation was undertaken as per the simulation study discussed previously.

The approximate posterior means and standard deviations of the VB distributions were all close to the posterior means and standard deviations obtained using MCMC (refer

to Table 4 and Figure 5). The regression coefficients are all positive and statistically significantly different from zero. From an examination of the predictive distributions of the PAO for the different species it is evident that the VB distributions can be used to obtain accurate approximations to the true predictive distributions of the PAO (refer to Figure 6 and Table 5). Notice that the accuracy statistics for all of the species considered were above 0.9.

Table 4. Parameter estimates of the single season occupancy models fitted using MLE, VB and MCMC.

		MLE		**VB**		**MCMC**		
Species	Covariate	Est	SE	Mean	Std	Mean	Std	MCSE
Black headed Heron	Int (detection)	0.412	0.123	0.413	0.113	0.408	0.124	<0.001
	nspp	0.381	0.128	0.381	0.123	0.386	0.128	<0.001
	Int (Occupancy)	1.136	0.255	1.131	0.231	1.182	0.269	0.002
	AETdivPETs	0.892	0.236	0.890	0.221	0.926	0.246	0.002
Egyptian Goose	Int (detection)	0.738	0.124	0.739	0.117	0.737	0.125	<0.001
	nspp	0.745	0.135	0.745	0.131	0.751	0.135	<0.001
	Int (Occupancy)	1.690	0.300	1.682	0.273	1.764	0.323	0.003
	AETdivPETs	0.866	0.258	0.861	0.242	0.906	0.273	0.002
Orange throated longclaw	Int (detection)	0.946	0.141	0.946	0.135	0.950	0.143	0.001
	nspp	0.607	0.159	0.607	0.154	0.618	0.160	0.001
	Int (Occupancy)	0.668	0.238	0.665	0.229	0.690	0.243	0.001
	AETdivPETs	1.410	0.268	1.406	0.258	1.462	0.279	0.002
White browed sparrow weaver	Int (detection)	0.372	0.132	0.374	0.120	0.370	0.133	0.001
	nspp	0.280	0.132	0.280	0.127	0.284	0.132	<0.001
	Int (Occupancy)	0.607	0.210	0.604	0.197	0.626	0.213	0.001
	AETdivPETs	0.556	0.205	0.556	0.196	0.569	0.210	0.001
Long tailed widow bird	Int (detection)	1.257	0.153	1.256	0.149	1.268	0.154	0.001
	nspp	0.727	0.168	0.726	0.166	0.738	0.169	0.001
	Int (Occupancy)	0.746	0.245	0.743	0.241	0.763	0.251	0.002
	AETdivPETs	1.633	0.293	1.628	0.288	1.693	0.302	0.002

The estimation results indicate that the VB method accurately estimates the posterior means of all of the parameters while the posterior variances are marginally underestimated.

2.3 Discussion

We developed two new methods of approximating the posterior distribution of the parameters of a Bayesian single season occupancy model that use logistic link functions. The first method uses a Laplace approximation of the VB optimal distributions while the second method utilizes the tangent based method of Jaakkola and Jordan (2000). Based on the simulation studies it was found that the Laplace approximation method performed well under most conditions considered. We believe that the approximation results obtained

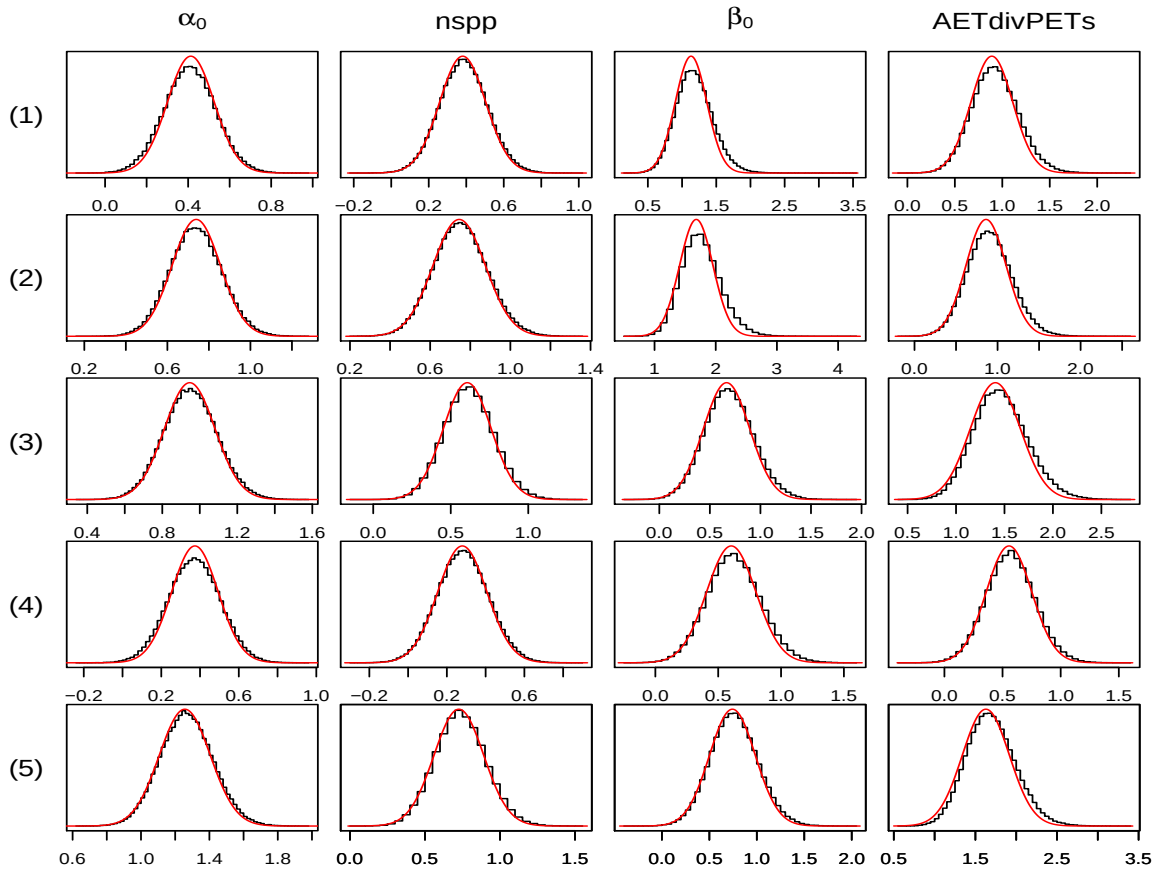

Figure 5. A comparison between the VB distributions (solid line) and the posterior distributions obtained using MCMC (the histogram) for the regression parameters of the detection and occupancy process for the different bird species (denoted as (1) = Black-headed heron, (2) = Egyptian goose, (3) = Orange-throated longclaw, (4) = White-browed sparrow-weaver and (5) = Long-tailed widowbird).

Table 5. Summary statistics of the posterior predictive distributions of the proportion of sites occupied using the MCMC method and the VB method for the five bird species considered.

Species	Method	Mean	Std	2.5%	97.5%
Black-headed	MCMC	89.7	2.05	86	94
Heron	VB	89.3	1.64	87	93
Egyptian	MCMC	100.2	1.78	97	104
goose	VB	99.9	1.47	97	103
Orange-throated	MCMC	77.4	1.23	76	80
longclaw	VB	77.2	1.05	76	80
White-browed	MCMC	78.2	1.99	75	83
sparrow-weaver	VB	77.9	1.63	75	81
Long-tailed	MCMC	78.5	0.74	78	80
widowbird	VB	78.5	0.67	78	80

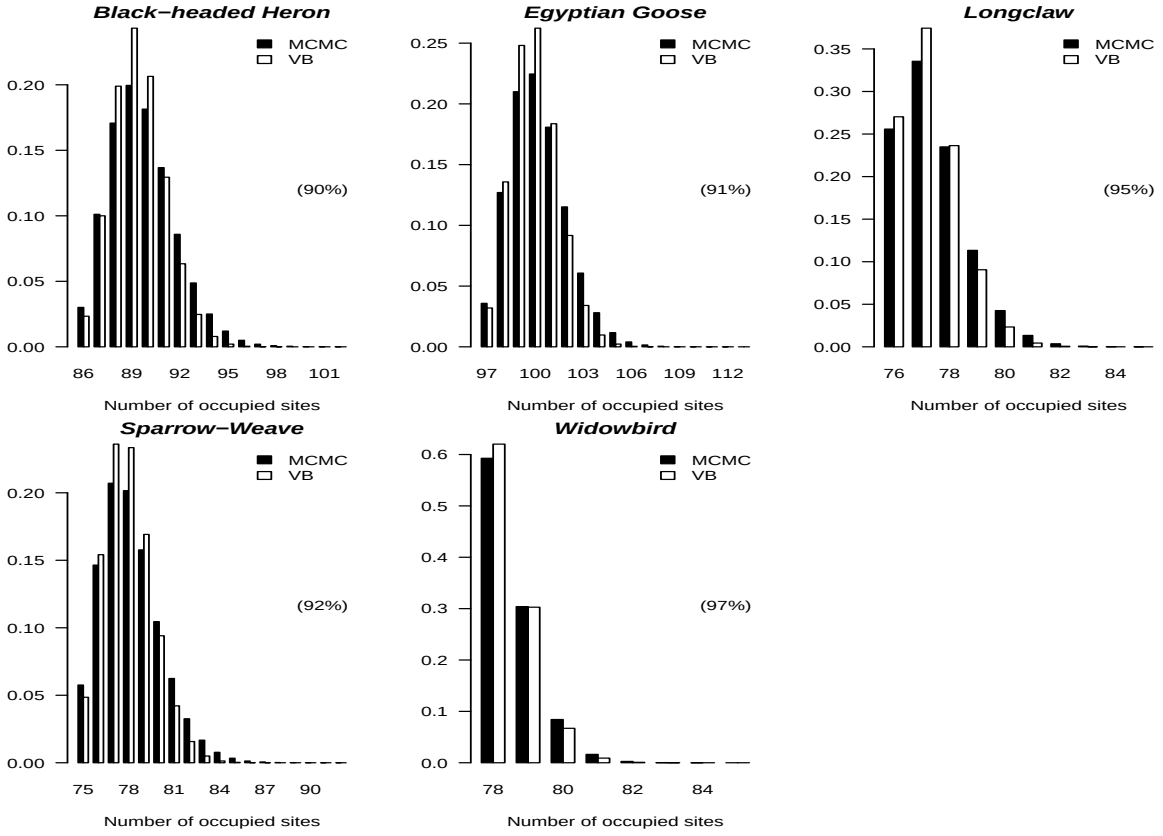

Figure 6. Predictive distribution of the proportions of occupied sites using the VB Laplace method and the MCMC method for the different bird species. The accuracy statistics ($acc(x)$) are displayed in brackets. The $acc(x)$ measure lies between 0 and 1 with a value of 1 indicating a perfect approximation and a value close to 0 indicating a poor approximation by the variational distribution to the true posterior distribution.

using the probit link function would be similar to those obtained using the tangent based method and thus did not explicitly consider this link function here. The methods have laid the groundwork that would enable VB methods to be applied to more complicated occupancy models and are currently the focus of ongoing research.

One big advantage of the methods developed here is the fact that they could be applied to cases where the researcher has informative prior information and might not want to rely on the use of the MLE method. In that situation, MCMC methods were so far the only methods available for fitting occupancy models in a Bayesian analysis. However, for big data sets, MCMC methods can be too slow to be useful. The code used to implement the methods is available in R and was at least 100 times faster than running MCMC using jagsUI **in our example**.

Simulations showed that when uninformative prior distributions were used, in general, the Laplace method attains very similar frequentist coverage probabilities to those obtained by the MLE method when the number of sampling occasions is at least three. We advise that the approximate methods could be used when the detection probability is at least 0.5

and there are at least three sampling occasions.

A further advantage of the methods developed here is the ease with which one can approximate the predictive distribution of the proportion of area occupied. Our simulation results showed that the Laplace approximate method can be used to obtain approximate distributions of the PAO. For scenarios where the detection probabilities are relatively low and the number of sites visits are small ($K = 2$) we found that the approximate methods slightly underestimate the upper bound of the PAO. The differences between the true predictive distribution and the approximate one is however very small for $K \geq 3$.

In both of the methods considered the approximate distributions derived are both multivariate Gaussian. When the sample size is particularly small, the number of sampling occasions is low (possibly one or two) or when the detection probability is low (less than 0.3) we have found that the posterior distributions of the parameters of the model are often skewed (particularly the occupancy covariate parameters). In these cases the approximate methods do not work well. Future work could entail the use of non-symmetric distributions similar to that proposed by Ormerod (2011).

Appendix

2.A Some matrix notation used in text.

Below we list some notation and distribution theory used in the main part of the text.

If $\boldsymbol{A}$ is the $b \times b$ matrix with entries denoted as a_{ij} $i = 1, \ldots, b,\ j = 1, \ldots, b$, $\boldsymbol{B} = \text{diag}(\boldsymbol{A})$ is the diagonal matrix with diagonal elements $b_{ii} = a_{ii}$ and off-diagonal elements all equal to zero. The identity diag($\boldsymbol{A}$) denotes the $b \times 1$ vector containing the diagonal elements of the matrix $\boldsymbol{A}$.

The trace of a matrix is defined as the sum of the diagonal elements of a matrix such that $\text{tr}(\boldsymbol{A}) = \sum_i a_{ii}$.

The transpose of the matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{A}^T$.

Define $b(x)$ as some real valued function such that $b'(x)$ and $b''(x)$ are the 1st and 2nd derivative of $b(x)$ respectively. Let $\boldsymbol{l}$ be a column vector such that $b(\boldsymbol{l})$ is a column vector where the i^{th} element is $b(l_i)$.

$\boldsymbol{A} \odot \boldsymbol{B}$ denotes the Hadamard product. This operation performs element-wise multiplication of the elements in $\boldsymbol{A}$ and $\boldsymbol{B}$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable matrices.

If $\boldsymbol{X}_p \sim N(\boldsymbol{\mu}\ \boldsymbol{\Sigma})$ then for some matrix $\boldsymbol{\Gamma}$, $\mathbb{E}(\boldsymbol{X}^T\boldsymbol{\Gamma}\boldsymbol{X}) = \text{tr}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{\Gamma}\boldsymbol{\mu}$. The entropy of the Gaussian distribution is defined as $\int f(\boldsymbol{x}) \ln f(\boldsymbol{x}) d\boldsymbol{x} = -\frac{p}{2}(\ln(2\pi) + 1) - \frac{1}{2}\log|\boldsymbol{\Sigma}|$.

2.B Lower bound to the joint likelihood and the VB distributions.

If we assume that the prior distribution for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are multivariate Gaussian distributions (denoted as $\pi(\boldsymbol{\alpha}, \boldsymbol{\beta})$) with parameters $\boldsymbol{\mu}^0_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}^0_{\boldsymbol{\alpha}}$ and $\boldsymbol{\mu}^0_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}^0_{\boldsymbol{\beta}}$ respectively then;

$$p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \prod_{i=1}^{n} o_i^{z_i} (1 - o_i)^{1 - z_i} \prod_{j=1}^{K_j} (z_i d_{i,j})^{y_{i,j}} (1 - z_i d_{i,j})^{1 - y_{i,j}}$$

where we assume that we visit n sites $\boldsymbol{K} = [K_1, K_2, \ldots, K_n]^T$ times respectively. Let

$b(x) = \ln(1 + \exp(x))$. $\ln \underline{p} = \ln p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ is equal to

$$\begin{aligned}\ln \underline{p} &= \sum_{\{i:z_i=1\}} \sum_{j=1}^{K_j} \left(y_{i,j} \boldsymbol{w}_{i,j}^T \boldsymbol{\alpha} - b\left(\boldsymbol{w}_{i,j}^T \boldsymbol{\alpha} \right) \right) + \boldsymbol{z}^T \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &= \boldsymbol{y}^T \mathrm{diag}(\tilde{\boldsymbol{z}}) \boldsymbol{W}\boldsymbol{\alpha} - \tilde{\boldsymbol{z}}^T b\left(\boldsymbol{W}\boldsymbol{\alpha}\right) + \boldsymbol{z}^T \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}). \end{aligned} \tag{2.7}$$

Equation (2.7) follows after some simplification and since when $z_i = 0$ we have $\boldsymbol{y}_i = \mathbf{0}$. Here $\tilde{\boldsymbol{z}} = [z_1 \mathbf{1}_{K_1}^T, \ldots, z_n \mathbf{1}_{K_n}^T]^T$.

A lower bound on the marginal log-likelihood can be obtained using

$$\ln p(\boldsymbol{y}) \geq \sum_{\boldsymbol{z}} \int q(\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \ln \left(\frac{p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{q(\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})} \right) d\boldsymbol{\alpha} d\boldsymbol{\beta}, \tag{2.8}$$

where $q(\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = q(\boldsymbol{z}) q(\boldsymbol{\alpha}) q(\boldsymbol{\beta})$. VB methods approximate the joint distribution of $\pi(\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta} | \boldsymbol{y})$ by maximising the lower bound of the marginal log-likelihood and providing update equations for $q(\boldsymbol{z})$, $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ in turn. Note that the dependence on $\boldsymbol{y}$ has been excluded from the notation for the variational distributions in order to save space.

Note that the variational posterior distribution of $\boldsymbol{\alpha}$ follows since

$$\begin{aligned}\ln q(\boldsymbol{\alpha}) &\propto \mathbb{E}_{\boldsymbol{\beta}, \boldsymbol{z}} \left(\ln \underline{p} \right) \\ &\propto \mathbb{E}_{\boldsymbol{\beta}, \boldsymbol{z}} \left(\boldsymbol{y}^T \mathrm{diag}(\tilde{\boldsymbol{p}}) \boldsymbol{W}\boldsymbol{\alpha} - \tilde{\boldsymbol{p}}^T b\left(\boldsymbol{W}\boldsymbol{\alpha}\right) + \boldsymbol{z}^T \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right) \\ &\propto \boldsymbol{y}^T \mathrm{diag}(\tilde{\boldsymbol{p}}) \boldsymbol{W}\boldsymbol{\alpha} - \tilde{\boldsymbol{p}}^T b\left(\boldsymbol{W}\boldsymbol{\alpha}\right) + \ln \pi(\boldsymbol{\alpha}). \end{aligned}$$

Similarly

$$\begin{aligned}\ln q(\boldsymbol{\beta}) &\propto \mathbb{E}_{\boldsymbol{\alpha}, \boldsymbol{z}} \left(\ln \underline{p} \right) \\ &\propto \mathbb{E}_{\boldsymbol{\alpha}, \boldsymbol{z}} \left(\boldsymbol{y}^T \mathrm{diag}(\tilde{\boldsymbol{p}}) \boldsymbol{W}\boldsymbol{\alpha} - \tilde{\boldsymbol{p}}^T b\left(\boldsymbol{W}\boldsymbol{\alpha}\right) + \boldsymbol{p}^T \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \right) \\ &\propto \boldsymbol{p}^T \boldsymbol{X}\boldsymbol{\beta} - \mathbf{1}_n^T b\left(\boldsymbol{X}\boldsymbol{\beta}\right) + \ln \pi(\boldsymbol{\beta}). \end{aligned}$$

2.C A tangent based method.

The integrations required to obtain the lower bound of the marginal log-likelihood are non-trivial and no closed form equations for $q(\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ exists. Additional variational parameters $\boldsymbol{v} = [\boldsymbol{a}^T, \boldsymbol{b}^T]^T$ are now introduced such that a lower bound of the marginal log-likelihood is found using

$$\ln p(\boldsymbol{y}) \geq \sum_{\boldsymbol{z}} \int q_{\boldsymbol{v}}(\boldsymbol{z}) q_{\boldsymbol{v}}(\boldsymbol{\alpha}) q_{\boldsymbol{v}}(\boldsymbol{\beta}) \ln \left(\frac{p_{\boldsymbol{v}}(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})}{q_{\boldsymbol{v}}(\boldsymbol{z}) q_{\boldsymbol{v}}(\boldsymbol{\alpha}) q_{\boldsymbol{v}}(\boldsymbol{\beta})} \right) d\boldsymbol{\alpha} d\boldsymbol{\beta}. \tag{2.9}$$

$\boldsymbol{a} = [\underline{\boldsymbol{a}}_1^T, \ldots, \underline{\boldsymbol{a}}_n^T]^T$ where each of the $\underline{\boldsymbol{a}}_i$ vectors are of length K_i while $\boldsymbol{b}$ is of length n. $p_{\boldsymbol{v}}(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ should be viewed as a lower bound for $p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})$ while $q_{\boldsymbol{v}}(\boldsymbol{z})$, $q_{\boldsymbol{v}}(\boldsymbol{\alpha})$ and $q_{\boldsymbol{v}}(\boldsymbol{\beta})$ indicates that the distributions of $\boldsymbol{z}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are functions of $\boldsymbol{v}$ which could have different functional forms.

The function $f(x) = -\ln(1+e^x)$ can be presented as the maximum of a family of parabolas (Jaakkola and Jordan, 2000) where

$$\begin{aligned} -\ln(1+e^x) &= \max_{\xi \in \Re} \left(A(\xi)x^2 - \frac{1}{2}x + C(\xi) \right) \text{ for all } x \in \Re, \\ A(\xi) &= -\tanh\left(\frac{1}{2}\xi\right)/(4\xi), \\ C(\xi) &= \frac{1}{2}\xi - \ln(1+e^{\xi}) + \xi \tanh\left(\frac{1}{2}\xi\right)/4. \end{aligned}$$

A lower bound for the quantities $b(\boldsymbol{W}\boldsymbol{\alpha})$ and $b(\boldsymbol{X}\boldsymbol{\beta})$ are obtained by using the above result such that

$$\begin{aligned} \ln \underline{p} \;&\geq\; \boldsymbol{y}^T \mathrm{diag}(\tilde{\boldsymbol{z}}) \boldsymbol{W}\boldsymbol{\alpha} + \boldsymbol{z}^T \boldsymbol{X}\boldsymbol{\beta} + \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \tilde{\boldsymbol{z}}^T \left(A(\boldsymbol{a}) \odot (\boldsymbol{W}\boldsymbol{\alpha})^2 - \frac{1}{2}\boldsymbol{W}\boldsymbol{\alpha} + C(\boldsymbol{a}) \right) \\ &\quad + \mathbf{1}_n^T \left(A(\boldsymbol{b}) \odot (\boldsymbol{X}\boldsymbol{\beta})^2 - \frac{1}{2}\boldsymbol{X}\boldsymbol{\beta} + C(\boldsymbol{b}) \right) \\ &\geq\; -\frac{1}{2}\boldsymbol{\alpha}^T \boldsymbol{B}_1 \boldsymbol{\alpha} + \boldsymbol{B}_2 \boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{D}_1 \boldsymbol{\beta} + \boldsymbol{D}_2 \boldsymbol{\beta} + \tilde{\boldsymbol{z}}^T C(\boldsymbol{a}) + \mathbf{1}_n^T C(\boldsymbol{b}) + E. \end{aligned} \tag{2.10}$$

Note that

$$\begin{aligned} \tilde{\boldsymbol{z}}^T \left(A(\boldsymbol{a}) \odot (\boldsymbol{W}\boldsymbol{\alpha})^2 \right) &= \boldsymbol{\alpha}^T \boldsymbol{W}^T \mathrm{diag}(A(\boldsymbol{a}) \odot \tilde{\boldsymbol{z}}) \boldsymbol{W}\boldsymbol{\alpha} \\ 1_n^T \left(A(\boldsymbol{b}) \odot (\boldsymbol{X}\boldsymbol{\beta})^2 \right) &= \boldsymbol{\beta}^T \boldsymbol{X}^T \mathrm{diag}(A(\boldsymbol{b})) \boldsymbol{X}\boldsymbol{\beta} \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{B}_1 &= (\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1} - 2\boldsymbol{W}^T \mathrm{diag}(A(\boldsymbol{a}) \odot \tilde{\boldsymbol{p}}) \boldsymbol{W} \\ \boldsymbol{B}_2 &= \left(\tilde{\boldsymbol{P}}\boldsymbol{y} - \frac{1}{2}\tilde{\boldsymbol{p}} \right)^T \boldsymbol{W} + (\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0)^T (\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1} \\ \boldsymbol{D}_1 &= (\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1} - 2\boldsymbol{X}^T \mathrm{diag}(A(\boldsymbol{b})) \boldsymbol{X} \\ \boldsymbol{D}_2 &= \left(\boldsymbol{p} - \frac{1}{2}\mathbf{1}_n \right)^T \boldsymbol{X} + (\boldsymbol{\mu}_{\boldsymbol{\beta}}^0)^T (\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1} \\ E &= -\frac{1}{2} \left(\ln 2\pi(p+q) + \ln |\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0| + \ln |\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0| + (\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0)^T (\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1} \boldsymbol{\mu}_{\boldsymbol{\alpha}}^0 + (\boldsymbol{\mu}_{\boldsymbol{\beta}}^0)^T (\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}}^0 \right). \end{aligned}$$

From equation (2.10) it is apparent that both $q_{\boldsymbol{v}}(\boldsymbol{\alpha})$ and $q_{\boldsymbol{v}}(\boldsymbol{\beta})$ are multivariate Gaussian distributions; specifically $q_{\boldsymbol{v}}(\boldsymbol{\alpha}) \sim \mathcal{N}(\boldsymbol{B}_1^{-1}\boldsymbol{B}_2^T \;,\; \boldsymbol{B}_1^{-1})$ and $q_{\boldsymbol{v}}(\boldsymbol{\beta}) \sim \mathcal{N}(\boldsymbol{D}_1^{-1}\boldsymbol{D}_2^T \;,\; \boldsymbol{D}_1^{-1})$. Using various known matrix identities regarding multivariate Gaussian distributions it can

be shown that

$$c_i^{(T)} = \boldsymbol{x}_i^T \boldsymbol{\mu}_{\boldsymbol{\beta}} + \mathbf{1}_{K_i}^T C(\underline{\boldsymbol{a}}_i) - \frac{1}{2}\mathbf{1}_{K_i}^T \underline{\boldsymbol{w}}_i \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \operatorname{tr}(\boldsymbol{d}_i) \tag{2.11}$$

where $\boldsymbol{d}_i = \underline{\boldsymbol{w}}_i^T \operatorname{diag}(A(\underline{\boldsymbol{a}}_i))\underline{\boldsymbol{w}}_i \left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} + \boldsymbol{\mu}_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\right)$. One way of estimating the variational parameters is to numerically maximise the right hand side of equation (2.9) with respect to $\boldsymbol{K}$. This approach might be feasible although it is not attempted here. Similar to Jaakkola and Jordan (2000) we however devise a Tangent algorithm in order to estimate $\boldsymbol{v}$.

2.C.1 Estimation of the variational parameters - The E-step

Denote the 'new variational parameters' as $\boldsymbol{a}_{(\boldsymbol{N})}$ and $\boldsymbol{b}_{(\boldsymbol{N})}$ respectively. Further denote $\boldsymbol{v}^{(new)}$ as $\boldsymbol{N}$ and $\boldsymbol{v}^{(old)}$ as $\boldsymbol{O}$ for notational convenience below. Treating $\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}$ as the 'complete data', the E-step of the Tangent algorithm is found by calculating the conditional expectation of the right hand side of equation (2.10) which equals $Q(\boldsymbol{N}|\boldsymbol{O}) = T_{\boldsymbol{\alpha},\boldsymbol{N}} + T_{\boldsymbol{\beta},\boldsymbol{N}} + E$ where

$$\begin{aligned} T_{\boldsymbol{\alpha},\boldsymbol{N}} &= \operatorname{tr}\left(-\frac{1}{2}\boldsymbol{B}_{1,\boldsymbol{N}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{O}} + \boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{O}}\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{O}}^T\right)\right) + \boldsymbol{B}_2\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{O}} + \boldsymbol{B}_{3,\boldsymbol{N}} \\ T_{\boldsymbol{\beta},\boldsymbol{N}} &= \operatorname{tr}\left(-\frac{1}{2}\boldsymbol{D}_{1,\boldsymbol{N}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{O}} + \boldsymbol{\mu}_{\boldsymbol{\beta},\boldsymbol{O}}\boldsymbol{\mu}_{\boldsymbol{\beta},\boldsymbol{O}}^T\right)\right) + \boldsymbol{D}_2\boldsymbol{\mu}_{\boldsymbol{\beta},\boldsymbol{O}} + \boldsymbol{D}_{3,\boldsymbol{N}}. \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{B}_{3,N} &= \tilde{\boldsymbol{p}}^T C(\boldsymbol{a}_{\boldsymbol{N}}) - \frac{1}{2}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0)^T(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0) \\ \boldsymbol{D}_{3,\boldsymbol{N}} &= \mathbf{1}_n^T C(\boldsymbol{b}_{\boldsymbol{N}}) - \frac{1}{2}(\boldsymbol{\mu}_{\boldsymbol{\beta}}^0)^T(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\beta}}^0). \end{aligned}$$

Note that here the additional subscript $\boldsymbol{N}$ indicates the dependence on the 'new variational parameters' which are estimated in the M-step, while subscript $\boldsymbol{O}$ indicates the dependence on the 'old variational parameters'.

2.C.2 Estimation of the variational parameters - The M-step

Since $Q(\boldsymbol{N}|\boldsymbol{O})$ separates in two functions of which one only depends on $\boldsymbol{a}_{(\boldsymbol{N})}$ and the second only depends on $\boldsymbol{b}_{(\boldsymbol{N})}$, it can be shown that

$$(\boldsymbol{a}_{(\boldsymbol{N})})^2 = \operatorname{diag}\left(\boldsymbol{W}\left(\boldsymbol{\Sigma}_{\boldsymbol{\alpha},\boldsymbol{O}} + \boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{O}}\boldsymbol{\mu}_{\boldsymbol{\alpha},\boldsymbol{O}}^T\right)\boldsymbol{W}^T\right) \tag{2.12}$$

$$\left(\boldsymbol{b}_{(\boldsymbol{N})}\right)^2 = \operatorname{diag}\left(\boldsymbol{X}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta},\boldsymbol{O}} + \boldsymbol{\mu}_{\boldsymbol{\beta},\boldsymbol{O}}\boldsymbol{\mu}_{\boldsymbol{\beta},\boldsymbol{O}}^T\right)\boldsymbol{X}^T\right). \tag{2.13}$$

2.D Convergence calculations.

The VB method is undertaken by maximising the lower bound to the marginal log-likelihood,

$$\mathbb{E}_q(\ln \underline{p}) - \mathbb{E}_q(\ln q(\boldsymbol{\alpha}, \boldsymbol{\beta})) - \mathbb{E}_q(\ln q(\boldsymbol{z})).$$

The VB algorithm is started by setting initial starting values for the parameters of the VB distributions as well as the variational parameters and are updated as the progresses from iteration to iteration. Denote the value of the lower bound at iteration t as A_t. Both algorithms are stopped when $|A_t - A_{t-1}| \leq \epsilon$ where $\epsilon = 10^{-6}$.

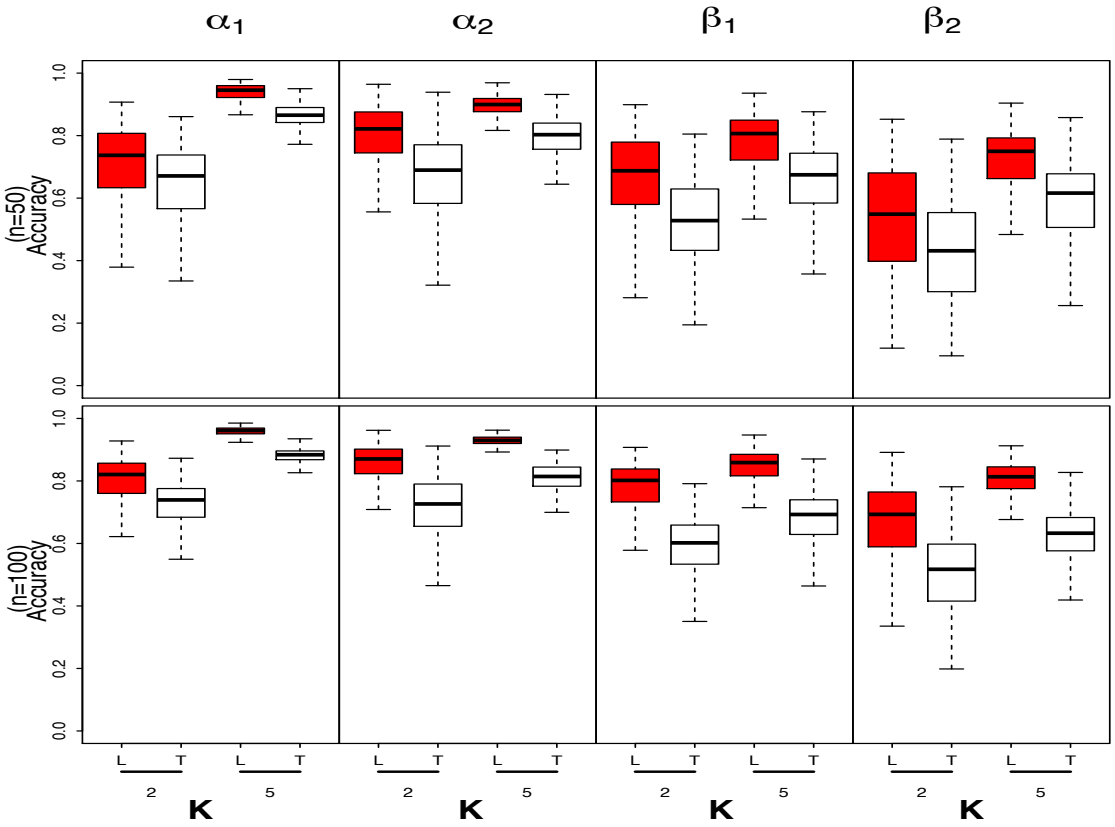

Figure 7. Box plots of the accuracy measurements for the model parameters associated with the Laplace (L-dark boxes) and Tangent (T-light boxes) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. The detection probability is approximately 0.5 while the occupancy probability is approximately 0.3.

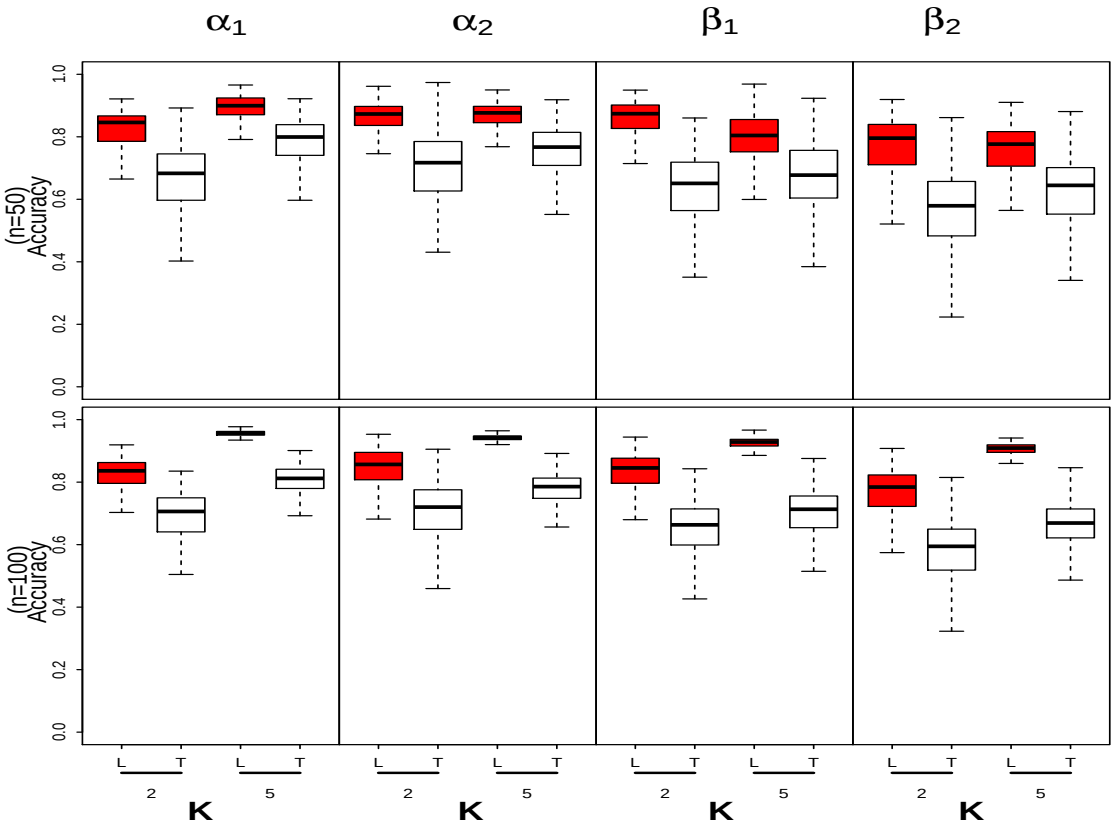

Figure 8. Box plots of the accuracy measurements for the model parameters associated with the Laplace (L-dark boxes) and Tangent (T-light boxes) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. The detection probability is approximately 0.7 while the occupancy probability is approximately 0.3.

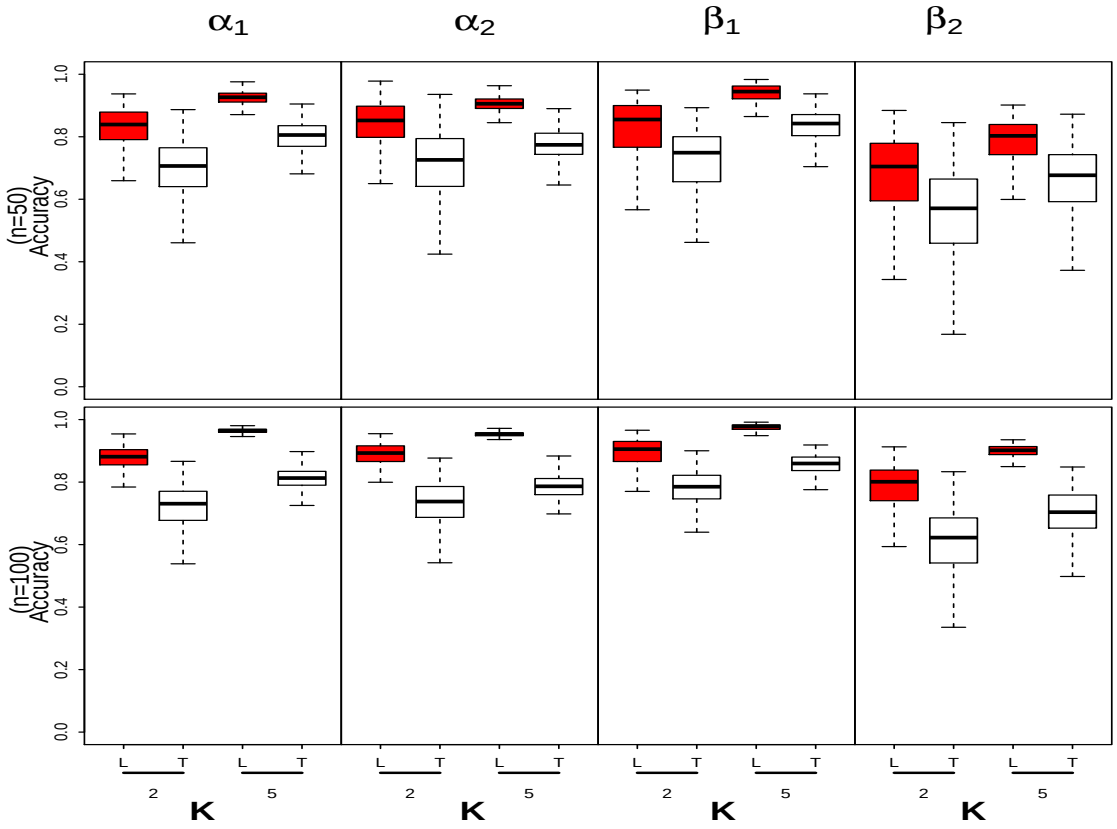

Figure 9. Box plots of the accuracy measurements for the model parameters associated with the Laplace (L-dark boxes) and Tangent (T-light boxes) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. The detection probability is approximately 0.7 while the occupancy probability is approximately 0.5.

Appendix

Additional material for Chapter 2

The following appendix consists of some of the supplementary material that accompanied **Clark A.E., Altwegg R., Ormerod J.T. (2016) A Variational Bayes Approach to the Analysis of Occupancy Models. PLoS ONE 11(2): e0148966. doi:10.1371/journal.pone.0148966.**

2.A VB analysis code

The R code used to undertake the analysis can be found at `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0148966#sec010`.

2.B VB Laplace approximation code explained

The VB Laplace approximation function and its usage is briefly explained below.

Function Call in R

The function call used to perform an analysis is as follows:

```
vb_model2_la(formula, design_mats, alpha_0, beta_0,  Sigma_alpha_0,
   Sigma_beta_0, LargeSample = FALSE, epsilon = 1e-05)
```

Named Arguments

The arguments of the function are as follows

- formula - Double right-hand side formula describing covariates of detection and occupancy in that order. e.g. Assume that the presence absence data is named y; the detection covariates is contained in a named list W (see below) and the occupancy covariates is stored X. Further suppose that the named lists are named W1, W2, W3 and X1 and X2 respectively. y $\sim$ W1 + W2 + W3 $\sim$ X1+ X2 would be one example of a suitable formula call. The function does not allow one to fit a model that only contains intercepts at the moment. This option will be included in future.

- design_mats - A named list generated by the call vb_Designs(W, X, y).

W is a named list of data frames of covariates that vary within sites. i.e. The data frames are of dimension $n \times J$ where each row is associated with a site and each column represents a site visit.

e.g. Suppose W contained three data frames W1, W2 and W3; W$W1[1,] = the covariate values for site 1 for all of the visits. Note that some of the entries might be 'NA' meaning that no visit took place at those occasions.

X is a named data frame that varies at site level.

y is an $n \times J$ matrix of the detection, non-detection data, where n is the number of sites, J is the maximum number of sampling periods per site.

NOTE: THE FUNCTION DOES NOT ALLOW THERE TO BE ANY MISSING VALUES IN THE COVARIATE MATRICES IF A SURVEY WAS UNDERTAKEN AT A PARTICULAR LOCATION!!!

- alpha_0 - Prior mean of the detection covariate coefficients. It is assumed that the detection covariate coefficients have the following prior distribution $\boldsymbol{\alpha} \sim \mathcal{N}$(alpha_0, Sigma_alpha_0). Here $\boldsymbol{\alpha}$ is viewed as a vector.

- beta_0 - Prior mean of the occurrence covariate coefficients. It is assumed that the occupancy covariate coefficients have the following prior distribution $\boldsymbol{\beta} \sim \mathcal{N}(\text{beta\_0}, \text{Sigma\_beta\_0})$. Here $\boldsymbol{\beta}$ is viewed as a vector.

- Sigma_alpha_0 - Prior covariance matrix of the detection covariate coefficients.

- Sigma_beta_0 - Prior covariance matrix of the occurrence covariate coefficients.

- LargeSample - LargeSample==TRUE - indicates that the number of sites is 'large' and that an approximation to $B(\mu, \sigma^2)$ is used instead of integrations (otherwise numerical integrations are performed).

- epsilon - Convergence measured relative to this quantity.

The values outputted by the function

- alpha - The VB estimate of the posterior mean vector of $\boldsymbol{\alpha}$. ($s \times 1$ vector)

- beta - The VB estimate of the posterior mean vector of $\boldsymbol{\beta}$. ($r \times 1$ vector)

- Sigma_alpha - The VB estimate of the posterior covariance matrix of the α vector. ($s \times s$ matrix)

- Sigma_beta - The VB estimate of the posterior covariance matrix of the $\boldsymbol{\beta}$ vector. ($r \times r$ matrix)

- occup_p - The VB estimate of the posterior occupancy probabilities at the sites considered. ($n \times 1$ vector)

- Log_mla - The lower bound of the log marginal log likelihood.

- Breakcounter - Breakcounter==1 if the number of iterations to perform the calculations are large. At the moment 'large' is viewed as 2000 iterations.

2.B.1 A small simulated data set

The following R code could be used to produce a small simulated data set that could be used to undertake the VB Laplace approximations.

```
#A simple example of how to construct y, X and W; the
#detection/nondetection data, site covariates and observation covariates
#-----------------------------------------------------------------------

require(MASS)
set.seed(1000)

beta.param = c(-1.85, 1.5, -0.5)
n = 5

#create 2 site covariates used to model occupancy
x1 = runif(n, -2,2)
x1 = (x1 - mean(x1)) / sd(x1)
x2 = runif(n, -5,5)
x2 = (x2 - mean(x2)) / sd(x2)
X = cbind(rep(1,n), x1, x2)
psi = as.vector(1/(1+exp(-X %*% beta.param))) ##logistic link function
    used
z = rbinom(n, size=1, prob=psi)

J = 3 #maximum number of surveys (some sites might have fewer visits)

#three observation covariates used to model the detection probs
alpha.param = c(-1.35, 1.0, 0.5, -.25)
w1 = runif(n*J, -5,5)
w1 = (w1 - mean(w1)) / sd(w1)
w2 = runif(n*J, -1,1)
w2 = (w2 - mean(w2)) / sd(w2)
w3 = runif(n*J, 0,5)
w3 = (w3 - mean(w3)) / sd(w3)
W = array(dim=c(n,J,4))
W[,,1] = 1
W[,,2] = w1
W[,,3] = w2
W[,,4] = w3

p = matrix(nrow=n, ncol=J)
```

```r
y = matrix(nrow=n, ncol=J)
for (j in 1:J)
{
   p[, j] = c(1/(1+exp(-W[,j,] %*% alpha.param)))
   y[, j] = rbinom(n, size=1, prob=z*p[, j])
}
#-------------------------------------------------------------------

#Now lets simulate the number of visits to each of the sites
#i.e. we need to set some of the y and W entries equal to
#NA
nvisits<-sample(1:J, n, replace=T)
empty.sites<-which(nvisits!= J)

for (i in 1:length(empty.sites))
{
   #adds NA to sites with visits less than J
   y[ empty.sites[i], (nvisits[empty.sites[i]]+1):J ] <- NA

   #adds NA to W entries with visits less than J
   W[ empty.sites[i], (nvisits[empty.sites[i]]+1):J, ] <- NA
}

#Note W[i,,] are the covariate values for site i
#each row is for a specific visit
#-------------------------------------------------------------------

#An nxJ matrix of the observed measured data,
#where n is the number of sites and J is the
#maximum number of observations per site.
Y.eg<-y
#-------------------------------------------------------------------

#siteCovs
#A data.frame of covariates that vary at the site level.
#This should have n rows and one column per covariate
X.eg=as.data.frame(cbind(x1,x2))
#-------------------------------------------------------------------

#obsCovs
#the obsCovs matrix is constructed as per the 'unmarked' package
#i.e. W.eg.l1 is a named list of data.frames of covariates that
#vary within sites.
#i.e. The dataframes are of dimension n by J
#where each row is associated with a site
#and each column represents a site visit.
#e.g. W.eg.l1$W1[1, ] = the covariate values for site 1 for all of the
#visits. Note that some of the entries might be 'NA'
#meaning that no visit took place at those occasions.

W1=matrix(NA,nrow=n, ncol=J)
W2=matrix(NA, nrow=n, ncol=J)
W3=matrix(NA, nrow=n, ncol=J)
for (i in 1:n)
{
   W1[i,]<- W[i,,2]
   W2[i,]<- W[i,,3]
   W3[i,]<- W[i,,4]
}
```

```
96
97  #colnames(W1)<-paste("W1.",1:J,sep="")
98  #colnames(W2)<-paste("W2.",1:J,sep="")
99  #colnames(W3)<-paste("W3.",1:J,sep="")
100
101 W.eg.l1<-list(W1=W1, W2=W2, W3=W3)
102 W.eg.l1
103 #-----------------------------------------------------------------
104
105 #An alternate way of 'viewing' the site covariates is as follows:
106 #Create a list element; one for each site, where the data
107 #for each site have been stacked one below the other either as
108 #a dataframe or as a matrix. e.g.
109 #W.eg.ls[[2]] is the data for site 2.
110
111 W.eg.l2=list(list())
112
113 for (i in 1:n)
114 {
115    if (nvisits[i]!=1)
116    {
117       dframe<-as.data.frame(W[i,1:nvisits[i],][,-1])
118    }else
119    {
120       dframe<-as.data.frame(matrix(W[i,1:nvisits[i],][-1],nrow=1))
121    }
122
123    names(dframe)<-c("w1","w2","w3")
124    W.eg.l2[[i]]<-dframe
125 }
126 #-----------------------------------------------------------------
127
128 #Two different ways of representing the observation covariates
129 W.eg.l1
130 W.eg.l2
131 #-----------------------------------------------------------------
132
133 #If the site covariates are provided as per W.eg.l1
134 #then we can construct W.eg as follows
135 #(here W.eg is the way in which 'vb_model2_la')
136 #creates the site covariate matrix W)
137 #We assume that all sites are visited at lest once
138 #although all might not be visited J times
139 #We further assume that there are no missing covariate
140 #values for those occasions sites are visited
141
142 W.temp<-NULL
143 n<-length(W.eg.l1)
144 for (i in 1:n)
145 {
146    W.temp<-cbind(W.temp, W.eg.l1[[i]])
147 }
148 W.temp
149
150 nvisits<-apply(W.eg.l1[[1]],1,function(x){length(na.omit(x))})
151 nvisits
152
153 W.eg<-NULL
154 for (i in 1:n)
```

```
{
   W.eg<-rbind(W.eg, matrix( c(na.omit(W.temp[i,])), nrow=nvisits[i]) )
}
W.eg
#-----------------------------------------------------------------

#If the site covariates are provided as per W.eg.l2
#then we can construct W.eg as follows
W.eg<-NULL
n <-length(W.eg.l2)
for (i in 1:n)
{
   W.eg<- rbind(W.eg, W.eg.l2[[i]])
}
W.eg
#-----------------------------------------------------------------

SimData<-list(y=Y.eg, X=X.eg, W.eg.l1=W.eg.l1, W.eg.l2=W.eg.l2, W_vb=W.
   eg)
```

The simulated data is stored in a list named SimData and it's contents are displayed below.

```
> SimData
$y
[,1] [,2] [,3]
[1,]    0   NA   NA
[2,]    0    1    0
[3,]    0   NA   NA
[4,]    0    1    0
[5,]    0    0   NA

$X
x1          x2
1 -0.5802233 -1.0949354
2  1.0468539  1.3166989
3 -1.3879419  0.7589495
4  0.7897816 -0.5628716
5  0.1315297 -0.4178415

$W.eg.l1
$W.eg.l1$W1
[,1]        [,2]        [,3]
[1,] -1.1475341          NA          NA
[2,]  0.3516124   1.4671572   0.2061681
[3,]  0.8607333          NA          NA
[4,] -1.1047705 -1.1515037 -1.0836617
[5,]  0.6777174   0.5505103          NA

$W.eg.l1$W2
[,1]        [,2]        [,3]
[1,]  0.08380091          NA          NA
[2,] -1.61591051 -1.7122869 -0.0359153
[3,]  0.06444337          NA          NA
[4,]  1.39241296   0.1392068   0.2994426
```

```
[5,] -0.10518068 -0.7375655          NA

$W.eg.l1$W3
[,1]        [,2]       [,3]
[1,]  0.6508060          NA         NA
[2,]  0.6021102  0.5024673 0.6489031
[3,] -0.0692492          NA         NA
[4,] -1.5268438 -0.2837397 1.4945716
[5,] -0.3372867  0.2584308         NA


$W.eg.l2
$W.eg.l2[[1]]
w1          w2        w3
1 -1.147534 0.08380091 0.650806

$W.eg.l2[[2]]
w1          w2         w3
1 0.3516124 -1.6159105 0.6021102
2 1.4671572 -1.7122869 0.5024673
3 0.2061681 -0.0359153 0.6489031

$W.eg.l2[[3]]
w1          w2          w3
1 0.8607333 0.06444337 -0.0692492

$W.eg.l2[[4]]
w1         w2          w3
1 -1.104771 1.3924130 -1.5268438
2 -1.151504 0.1392068 -0.2837397
3 -1.083662 0.2994426  1.4945716

$W.eg.l2[[5]]
w1          w2          w3
1 0.6777174 -0.1051807 -0.3372867
2 0.5505103 -0.7375655  0.2584308


$W_vb
w1           w2          w3
1   -1.1475341  0.08380091  0.6508060
2    0.3516124 -1.61591051  0.6021102
3    1.4671572 -1.71228685  0.5024673
4    0.2061681 -0.03591530  0.6489031
5    0.8607333  0.06444337 -0.0692492
6   -1.1047705  1.39241296 -1.5268438
7   -1.1515037  0.13920676 -0.2837397
8   -1.0836617  0.29944258  1.4945716
9    0.6777174 -0.10518068 -0.3372867
10   0.5505103 -0.73756554  0.2584308
```

2.B.2 A small example

The following R code could be used as an example of how to use the VB code in order to undertake a small analysis.

```
## Load the  data into your workspace
##-------------------------------------

#This data set is stored as a supplementary information document
#First download the file and then save it into your working directory
#before running the rest of the script
load("S2_Data.rda")

#Set Uninformative priors
#------------------------
#Coefficients in the detection model
alpha_0 <- matrix(0, ncol=1, nrow=4)

#Covariance matrix of the coefficients in the detection model
Sigma_alpha_0 <- diag(4)*1000

#Coefficients in the occupancy process
beta_0 <- matrix(0, ncol=1, nrow=3)

#Covariance matrix of the coefficients in the occupancy model
Sigma_beta_0 <- diag(3)*1000

#Construct the required matrices using vb_Designs
#-----------------------------------------------
#Ensure that the function 'vb_Designs' is stored in the workspace
#The function is included here if this was not done

vb_Designs<-function(W, X, y)
{
   #create the required 'response' and 'regressor matrices'
   #using all of the X and W data
   #the output is stored as a named list

   #create the Y matrix that will be used
   Y<-matrix(na.omit(matrix(t(y), ncol=1)))
   pres_abs <- apply(y,1,max,na.rm=T) #check if this will work for NA's

   #create the W matrix
   W.temp<-NULL
   nv<-length(W)

   for (i in 1:nv){W.temp<-cbind(W.temp, W[[i]])}

   nvisits<-apply(W[[1]],1,function(x){length(na.omit(x))})
   n<-length(nvisits)

   W.out<-NULL
   for (i in 1:n)
   {
      W.out<-rbind(W.out, matrix( c(na.omit(W.temp[i,])), nrow=nvisits[i
   ]) )
   }
   colnames(W.out)<-names(W)

   list(Y=as.data.frame(Y), X=as.data.frame(X), W=as.data.frame(W.out),
      Names=c( colnames(X), colnames(W.out)), nvisits=nvisits,
      pres_abs=pres_abs)
}

```

```
design_mats<-vb_Designs(W=SimData2$W.eg.l1, X=SimData2$X, y=SimData2$y)

#Here we use the large sample approximation and run the VB algorithm
#-------------------------------------------------------------------
#Assume that the formula used will be of the following form:
#formula1<- y~X1+X2~W1+W2+W3
#The occupancy model uses 2 covariates, X1 and X2; while
#the detection model uses 3 covariates W1, W2 and W3
#Intercepts are included in both models
#The function does not allow one to repress the intercept term

#ensure that the 'vb_model2_la' function is in the workspace
vb_fit<-vb_model2_la(y~X1+X2~W1+W2+W3,
   design_mats=design_mats,
   alpha_0=alpha_0, beta_0=beta_0,
   Sigma_alpha_0=Sigma_alpha_0, Sigma_beta_0=Sigma_beta_0,
   LargeSample=TRUE, epsilon=1e-5)

#The detection model parameters
vb_fit$alpha

#The occupancy model parameters
vb_fit$beta

#The respective covariance matrices
vb_fit$Sigma_alpha
vb_fit$Sigma_beta

#The approximate conditional ocupancy probabilities
plot(vb_fit$occup_p, ylab="Occupancy prob", xlab="Site number")
```

S1 Data

The R data file that contains the data used to undertake the analysis. See `https://doi.org/10.1371/journal.pone.0148966.s010`.

S2 Data

The R data file that contains the data used to explain how to use the VB Laplace approximation code. See `https://doi.org/10.1371/journal.pone.0148966.s011`.

Acknowledgements

This research was partially supported by an Australian Research Council Early Career Award DE130101670 (Ormerod), two South African National Research Foundation grants namely 81685 and 85802 (Altwegg), as well as a scholarship by the NRF (Clark). The financial assistance of the NRF towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

We hereby acknowledge that some of the computations were performed using facilities

provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za. We hereby thank the members of the Centre for Statistics in Ecology, Environment and Conservation for helpful comments in particular Greg Duckworth and Sanet Hugo. We also thank the three anonymous reviewers and the academic editor for their helpful comments, which greatly improved this manuscript.

We have no conflict of interest to declare.

Chapter 3

A Variational Bayes approach to the analysis of occupancy models - the probit case.

The R code used to undertake the analysis and simulation study can be found at `https://github.com/AllanClark/PhdAnalysisCode`.

Take note that the references for the article are all displayed at the end of the thesis.

'Allan Ernest Clark' wrote the paper and performed the mathematical development and analysis. 'Res Altwegg' and 'Allan Ernest Clark' reviewed the manuscript extensively. Both authors have read and approved the manuscript.

A Variational Bayes approach to the analysis of occupancy models - the probit case.

Allan E. Clark [1,2], Res Altwegg[1,2],

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.

Abstract

Clark et al. (2016) investigated the use of Variational Bayes (VB) to approximate the posterior distribution of the parameters of the single season occupancy (SSO) model when the regression effects are modelled using logistic link functions. In this paper, a method is developed to undertake the same task, when probit link functions are used instead, and the results were compared to those obtained when undertaking Markov Chain Monte Carlo (MCMC). This is done by developing two approximate methods, both utilising mean field Variational Bayes to obtain approximate posterior distributions of the parameters of the single season occupancy model. The first method assumes that the approximate distributions are Gaussian distributed, while the second method assumes that a linear mixture of Gaussian distributions is more appropriate. The effectiveness of the methods were investigated by means of a simulation study as well as the practical application to data from the 2nd Southern African Bird Atlas Project. It is shown that under certain conditions the variational distributions can provide accurate approximations to the true posterior distributions of the regression effects. A new method that combines VB and MCMC is suggested to obtain the posterior predictive distribution of the proportion of occupied sites.

3.1 Introduction

Bayesian analysis is a statistical paradigm in which a researcher's prior information regarding a subject area is combined with that of the data by means of a the likelihood function (Berger, 2013). Statistical inference is then undertaken by using posterior expectations of functions of the parameters of the model under investigation. Often, however, the posterior distributions are not of a known form thus making such calculations difficult. One way of undertaking posterior calculations is to utilise sampling methods such as *the bootstrap*, *the jackknife*, or one of the many Monte Carlo methods that have been developed (Efron, 1982). Alternatively one could use Markov Chain Monte Carlo methods (MCMC) such as the Gibbs sampler (Geman and Geman, 1984) or the Metropolis Hastings algorithm (Metropolis et al., 1953; Hastings, 1970).

Occupancy models constitute an important class of statistical model used to examine occupancy patterns, such as species distribution maps (MacKenzie et al., 2002). These methods are often undertaken using the Bayesian paradigm and numerous advances have been made in this area of research (Bailey et al., 2014). The single species occupancy (SSO) model has been widely investigated, often using classical methods (MacKenzie et al., 2002; Welsh et al., 2013; Guillera-Arroita et al., 2014; Moreno and Lele, 2010; Hutchinson et al., 2015). The first Bayesian formulation of this model was by Royle and Dorazio (2008) who used WinBUGS to undertake their analysis (Lunn et al., 2000). This led to many similar studies that utilized their code to answer various important biological questions (Hanks et al., 2011; Royle and Kery, 2007; Johnson et al., 2013). Later Dorazio and Rodriguez (2012) developed a Gibbs sampling algorithm for the model while Clark et al. (2016) investigated the use of Variational Bayes (Ormerod and Wand, 2010).

Variational Bayes (VB) is a method that approximates the posterior distributions of a models' parameters by minimizing the Kullback-Leibler divergence between the posterior distribution of the parameters of a model and the unknown approximating probability distribution (Kullback and Leibler, 1951). The method has origins in computer science and has extensively been used in the machine learning literature (MacKay, 1995; Ghahramani and Beal, 2000; Bishop, 2008). Occupancy models are hierarchical in nature with one aim being to relate the detection and occupancy processes of the model to biologically meaningful covariates. These relationships are modelled using either logistic or probit link functions. The methods developed below are used to approximate the posterior distributions of the parameters of the occupancy model obtained using the method developed by Dorazio and Rodriguez (2012).

Clark et al. (2016) found that the VB method could be used to approximate the posterior distribution of the parameters of the SSO model, as well as the predictive distribution of

the proportion of occupied sites (PAO), when logistic link functions are used to model the detection and occupancy probability regression coefficients. The method was particularly accurate when the true posterior distributions were in fact Gaussian distributed or close to Gaussian distributed. It was also found that under certain conditions the true posterior distributions were often skewed (with heavy tails to the right), in which case the variational approximations were not accurate and often underestimated the true posterior variance of the parameters of the model. In order to address this issue, Gershman et al. (2012) and Zobay (2014) investigated the use of Gaussian mixture distributions to approximate the posterior distributions obtained when undertaking logistic regression. Here we investigate the use of Gaussian mixture distributions to approximate the posterior distribution of the parameters of the SSO model focusing specifically on the probit link function. We do so since there exists a Gibbs sampling algorithm to sample from the parameters of the single season occupancy model when a probit link function is used to model the regression effects of the detection and occupancy processes (Dorazio and Rodriguez, 2012).

This paper commences with the derivation of a particular VB implementation of a SSO model that uses probit link functions and then proceeds to present the results of a short simulation study. Before concluding, with a summary of our findings we analyse detection-nondetection data of three bird species to illustrate the usefulness of the VB technique developed.

3.2 Materials and Methods

3.2.1 VB applied to single species occupancy models

One of the aims of a SSO model is to estimate the occupancy probability at the sites surveyed (n) during a study. Such an analysis is undertaken by collecting detection-nondetection data (denoted as $[y_{i,j}]$ for $i = 1, \ldots, n;\ j = 1, \ldots, K_i$ where K_i represents the number of surveys at site i) which entails visiting sites a number of times and recording whether or not the study species is observed at the site. $\boldsymbol{y}$ is defined to be the $N \times 1$[1] vector such that its elements $[\boldsymbol{y}_i]$ are $K_i \times 1$ vectors representing the detection-nondetection data of site i.

The SSO model can be represented using a hierarchical model where the occupancy and detection processes are both assumed to be Bernoulli random variables (Royle and Dorazio, 2008). The vector $\boldsymbol{z}$ represents the true species occupancy at the sites considered and is modelled as $z_i|\psi_i \sim \text{Bernoulli}(\psi_i)$, while the detection process is modelled as $y_{i,j}|z_i, d_{i,j} \sim \text{Bernoulli}(z_i d_{i,j})$ for all sites $i = 1, \ldots, n$ and all visits $j = 1, \ldots, K_i$. The occupancy probability ψ_i denotes the probability that the species occurs at site i while

[1] $N = \sum_{i=1}^{n} K_i$.

$d_{i,j}$ denotes the conditional probability of detecting the species during the j^{th} visit of site i given that the species is present at site i.

The notation presented here is the same as that used in Clark et al. (2016) except that the regression slope parameters (d_{ij} and ψ_i, $\forall(i,j)$) are modelled using a probit link function. Specifically, $d_{ij} = \Phi(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha})$ and $\psi_i = \Phi(\boldsymbol{x}_i^T\boldsymbol{\beta})$, such that the conditional likelihood of the data and the true occupancy variables is

$$p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\alpha},\boldsymbol{\beta}) = \prod_{i=1}^{n} \boldsymbol{\Phi}(\boldsymbol{x}_i^T\boldsymbol{\beta})^{z_i}(1-\boldsymbol{\Phi}(\boldsymbol{x}_i^T\boldsymbol{\beta}))^{1-z_i} \prod_{j=1}^{K_i}(z_i\boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}))^{y_{ij}}(1-z_i\boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha}))^{1-y_{ij}}.$$

In what follows we develop a VB approach to approximating the posterior distributions of the occupancy and detection regression parameters as well as the posterior distributions associated with the true occupancy variables, $\boldsymbol{z}$, where $q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z})$ represents the approximating distribution of $p(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z}|\boldsymbol{y})$. It is assumed that it can be factorized as $q(\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{z}) = q(\boldsymbol{\alpha})q(\boldsymbol{\beta})\prod_i q(z_i)$. Under this restriction $q(\boldsymbol{\alpha},\boldsymbol{\beta})$ can be factorized into two separate factors $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ with

$$q(\boldsymbol{\alpha}) \propto \exp(\boldsymbol{y}^T\tilde{\boldsymbol{P}}(a(\boldsymbol{W}\boldsymbol{\alpha}) - b(\boldsymbol{W}\boldsymbol{\alpha})) + \mathbf{1}_N^T\tilde{\boldsymbol{P}}b(\boldsymbol{W}\boldsymbol{\alpha}) + \ln\pi(\boldsymbol{\alpha})) \tag{3.1}$$

$$q(\boldsymbol{\beta}) \propto \exp(\boldsymbol{p}^T(a(\boldsymbol{X}\boldsymbol{\beta}) - b(\boldsymbol{X}\boldsymbol{\beta})) + \mathbf{1}_n^T b(\boldsymbol{X}\boldsymbol{\beta}) + \ln\pi(\boldsymbol{\beta})), \tag{3.2}$$

where $\tilde{\boldsymbol{p}} = \mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{\beta})}(\tilde{\boldsymbol{z}}|\boldsymbol{y})$, $\tilde{\boldsymbol{P}} = \text{diag}(\tilde{\boldsymbol{p}})$, $\boldsymbol{p} = \mathbb{E}_{q(\boldsymbol{\alpha},\boldsymbol{\beta})}(\boldsymbol{z}|\boldsymbol{y})$, $a(x) = \ln(\boldsymbol{\Phi}(x))$ and $b(x) = \ln(1-\boldsymbol{\Phi}(x))$. Vector evaluations using a general function $f(x)$ result in $f(\boldsymbol{X}) = [f(x_1),\ldots,f(x_n)]^T$. Here $\tilde{\boldsymbol{z}} = \left[z_1\mathbf{1}_{K_1}^T,\ldots,z_n\mathbf{1}_{K_n}^T\right]^T$.

The normalization constant of $q(\boldsymbol{\alpha},\boldsymbol{\beta})$ is not known analytically and thus $q(\boldsymbol{\alpha})$ and $q(\boldsymbol{\beta})$ are not of a known type and we thus approximate the posterior distribution of $(\boldsymbol{\alpha},\boldsymbol{\beta})$ using two methods. In the first method, multivariate Gaussian distributions (denoted as G) are used as the approximating distributions, whereas the second method uses linear Gaussian mixture distributions. In an occupancy modelling context, Clark et al. (2016) found that the use of multivariate Gaussian distributions results in the underestimation of the variance of the posterior distributions of the parameters of a SSO model (when logistic link functions are used to model the regression effects) and it is expected that the use of linear Gaussian mixture distributions would reduce this shortcoming. Appendix 3.A provides an introduction to linear Gaussian mixture distributions (denoted as GM).

In the second case we assume that

$$q(\boldsymbol{\alpha}) \sim \sum_{k=1}^{n_\alpha} w_{\boldsymbol{\alpha}}^{(k)}\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{(k)},\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(k)}) \text{ and}$$

$$q(\boldsymbol{\beta}) \sim \sum_{l=1}^{n_\beta} \boldsymbol{w}_{\boldsymbol{\beta}}^{(l)}\mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}^{(l)},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(l)})$$

where the weight parameters, $w_{\boldsymbol{\alpha}}^{(i)}$ and $w_{\boldsymbol{\beta}}^{(i)}$, and the Gaussian parameters, $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{(i)}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(i)}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}^{(i)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(k)}$ (for $i = k, l$ with $k = 1, \ldots, n_\alpha$ and $l = 1, \ldots, n_\beta$ respectively), are estimated using detection-nondetection data. The estimation is undertaken using the R package *iterlap*, which approximates a joint posterior distribution by means of a GM distribution (Bornkamp and Bornkamp, 2012). The elements of the GM distribution are constructed by initially approximating the posterior distribution (denoted in general as $\pi(\boldsymbol{\theta}|\boldsymbol{x})$), using a Laplace approximation where the approximate distribution is denoted as $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{x})$. A residual function is then obtained as $r(\boldsymbol{\theta}|\boldsymbol{x}) = \pi(\boldsymbol{\theta}|\boldsymbol{x}) - \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{x})$ which is also approximated by means of a Laplace approximation. The updated approximation of the posterior distribution is then defined as a linear combination of the the previous approximation and the approximation to the residual function. These steps are repeated until the approximation to the posterior distribution does not change. The number of components of each linear Gaussian mixture distribution (n_α and n_β) are not determined by the user. Bornkamp (2011) describes this procedure in detail.

Once we have obtained approximations to $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ it then follows that the q-densities, $q(z_i|\boldsymbol{y}_i = \mathbf{0})\ \forall i$, are Bernoulli distributed with success probability $(1 + \exp(-c_i))^{-1}$ where

$$c_i = \mathbb{E}_{\boldsymbol{\beta}}\left(a(\boldsymbol{x}_i^T\boldsymbol{\beta}) - b(\boldsymbol{x}_i^T\boldsymbol{\beta})\right) + \mathbf{1}_{K_i}^T \mathbb{E}_{\boldsymbol{\alpha}}\left(b(\boldsymbol{w}_{ij}^T\boldsymbol{\alpha})\right). \tag{3.3}$$

Note that there are no closed form expressions for the expectations in (3.3) although they can easily be obtained using numerical quadrature methods (Golub and Welsch, 1969). The parameters of the variational distributions are all dependent on one another and can be computed using an iterative scheme such as that given in Algorithm 3.

Algorithm 3. Iterative scheme for obtaining the parameters of the optimal density of $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using the GM distribution approximation.

1. Assume that $q(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be approximated using a multivariate Gaussian distribution.
   - Initialize the following matrices $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$.
   - Calculate conditional occupancy probabilities for all sites where $\boldsymbol{y}_i = \mathbf{0}$ using (3.3).
   - Obtain an initial estimate of the marginal log-likelihood.
2. Cycle:
   - Approximate $\ln q(\boldsymbol{\alpha})$ using a GM distribution to obtain $w_{\boldsymbol{\alpha}}^{(k)}$, $\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{(k)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{(k)}\ \forall k$.
   - Approximate $\ln q(\boldsymbol{\beta})$ using a GM distribution to obtain $w_{\boldsymbol{\beta}}^{(k)}$, $\boldsymbol{\mu}_{\boldsymbol{\beta}}^{(k)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{(k)}\ \forall k$.
   - Calculate conditional occupancy probabilities for all sites where $\boldsymbol{y}_i = \mathbf{0}$ using (3.3).

   until the change in the marginal log-likelihood becomes negligible ($\leq 10^{-6}$).

3.2.2 Simulation Study

In the following simulation study we investigate the accuracy of the VB approximations to the posterior distributions of the regression effects of the detection and occupancy

processes, as well as the predictive distribution of proportion of occupied sites (defined as PAO $= \frac{1}{n}\sum_{i=1}^{n} z_i$) obtained through MCMC. The details of the simulation study undertaken are similar to section 1.3 of Clark et al. (2016) and we *thus only repeat some of the details here.*

The accuracy of the VB approximations relative to the posterior distribution obtained through MCMC is assessed by calculating $\text{acc}(x) = 1 - \frac{1}{2}\int |q(x) - q_{MCMC}(x)|dx$. The $acc(x)$ measure lies between 0 and 1 with a value of 1 indicating a perfect approximation and a value close to 0 indicating a poor approximation by the variational distribution to the true posterior distribution.

In this study we consider 32 simulation settings. The number of sites (n) are set to 50 and 100 while the number of visits to each site (K) are set to $2, 3, 4$ and 5 respectively. The selected regression parameter values ensure an approximate average detection and occupancy probability among the sites of (0.5, 0.3), (0.7, 0.3), (0.5, 0.5) and (0.7, 0.5) respectively. The following prior distributions were used for the parameters of the SSO model: $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, 1000\boldsymbol{I}_2)$ and $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 1000\boldsymbol{I}_3)$.

Each simulation setting was replicated between 400 and 500 times[2]. All calculations were undertaken using R 3.4.0 (R Core Team, 2014). Numerical optimizations were performed using the BFGS method of the R function *optim*; numerical integrations were performed using the *gauss.hermite* function in the *spatstat* R package using *order=25* (Baddeley and Turner, 2005); MCMC sampling was undertaken by amending the R code found in Dorazio and Rodriguez (2012), while all variational approximations were performed using the authors' code which used the *iterlap* R package to fit GM distributions where needed. 100 000 posterior samples were obtained for each MCMC simulation, although the first 50 000 samples were discarded. Various convergence tests were undertaken to assess that the MCMC chains had converged (Smith, 2007). The posterior samples were not thinned (Link and Eaton, 2012).

3.3 Results

3.3.1 Simulation Results

Based on the accuracy calculations across the replicate data sets, the GM method is better than G at approximating the marginal posterior distributions of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ when comparisons are based on the mean/median accuracy measure for these parameters. As an example of these simulation results, consider the following two scenarios where the estimated mean detection and occupancy probability combinations across all sites and

[2]Replicates in which the estimation routine does not converge are not included in the analysis.

number of surveys are (0.5, 0.3); and (0.7, 0.3) respectively (as illustrated in Figure 10).

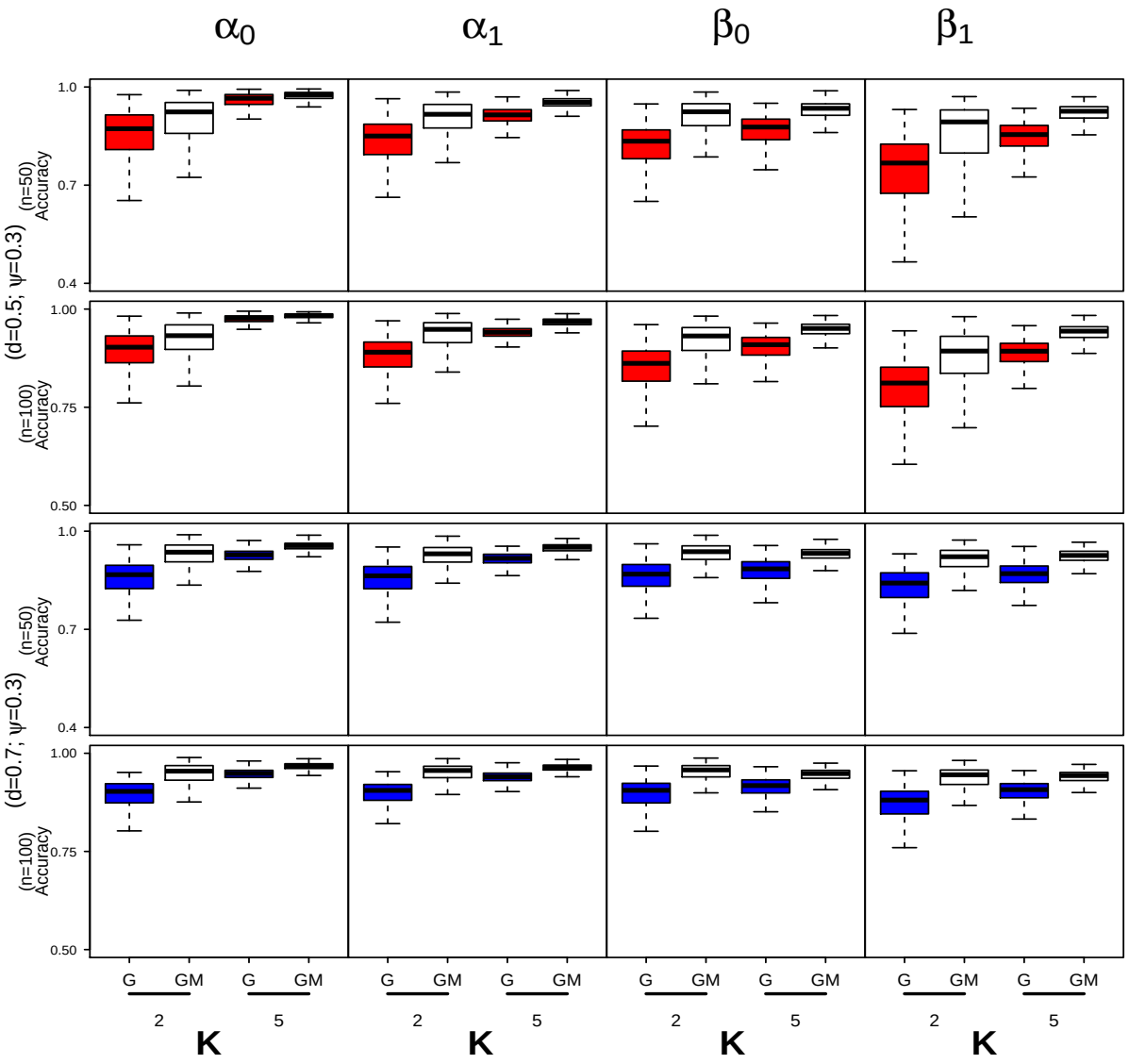

Figure 10. Boxplots of the accuracy measurements for the model parameters associated with the Gaussian (G) and Gaussian Mixture (GM) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. Two detection and occupancy probability scenarios are considered here namely $(d = 0.5,\ \psi = 0.3)$ and $(d = 0.7,\ \psi = 0.3)$.

In general, the posterior approximations for the detection regression parameters were quite good, with median accuracy statistics greater than 0.85. This was true even when the number of surveys are small $(K = 2, 3)$. The accuracy statistics for the detection regression parameters increased as n and K increased, with median accuracy statistics in excess of 0.95 when $K \geq 3$. Similar comments can be made regarding the accuracy of the posterior approximations for the occupancy regression parameters.

In general, accuracies increased with an increase in n and K although the rate of increased in the accuracy statistics for the occupancy regression parameters appears to be slower

than those observed for the detection regression parameters. The boxplots of the accuracy statistics associated with the different methods for the remaining cases can be found in Appendix 3.C (refer to Figure 14). We observed that the accuracy statistics for both methods improved with the increase in ψ.

Nonparametric t-tests were performed to assess whether the mean percentage of overlap for the Gaussian method is equal to the mean percentage of overlap for the GM method. The null hypothesis (equal means) is rejected for all cases which confirms what one observes in Figure 10.

We also found that the accuracy of the approximate predictive distributions of PAO improves as K increases (refer to Figure 11) and this observation is consistent across all of the scenarios considered. From an examination of the posterior predictive distributions (not shown here) it is evident that the VB predictive distributions are lighter tailed than the MCMC predictive distributions, although this effect is reduced for $K \geq 3$.

The five-number summary of the correlation coefficient between the occupancy probabilities obtained for methods G and GM (not shown) indicate that the occupancy probabilities obtained for the two methods are essentially the same. The boxplots of the accuracy measure for the predictive distribution of PAO for the two methods would therefore also be very similar (refer to Figure 11). These results were obtained since the posterior occupancy probabilities (obtained using equation (3.3)) are functions of the expectations of Gaussian and Gaussian mixture random variables which in these cases had very similar central tendencies. It can easily be shown that $z_i|\boldsymbol{\alpha}, \boldsymbol{\beta}$ is a Bernoulli random variable with success probability

$$s_i = \frac{\psi_i \prod_j (1 - p_{ij})}{1 - \psi_i + \psi_i \prod_j (1 - p_{ij})} \tag{3.4}$$

for sites where the species under investigation has not been seen (Dorazio and Rodriguez, 2012). An alternate method of obtaining the PAO is to sample from the VB distributions of the regression effects and the conditional distribution of z_i (using equation (3.4)) and in turn, to obtain posterior draws of the PAO. An application of this hybrid method which combines VB and MCMC to obtain a posterior estimate of the PAO can be found in Section 3.3.2 below.

3.3.2 Application using South African detection-nondetection data

Similarly to Clark et al. (2016), detection-nondetection data was extracted from the 2$^{\text{nd}}$ Southern African Bird Atlas Project database (Harebottle et al., 2007). This was done to compare the performance of three different methods for fitting a SSO model. The data was collected by citizen scientists during 2016 using 5-minute latitude $\times$ 5-minute longitude

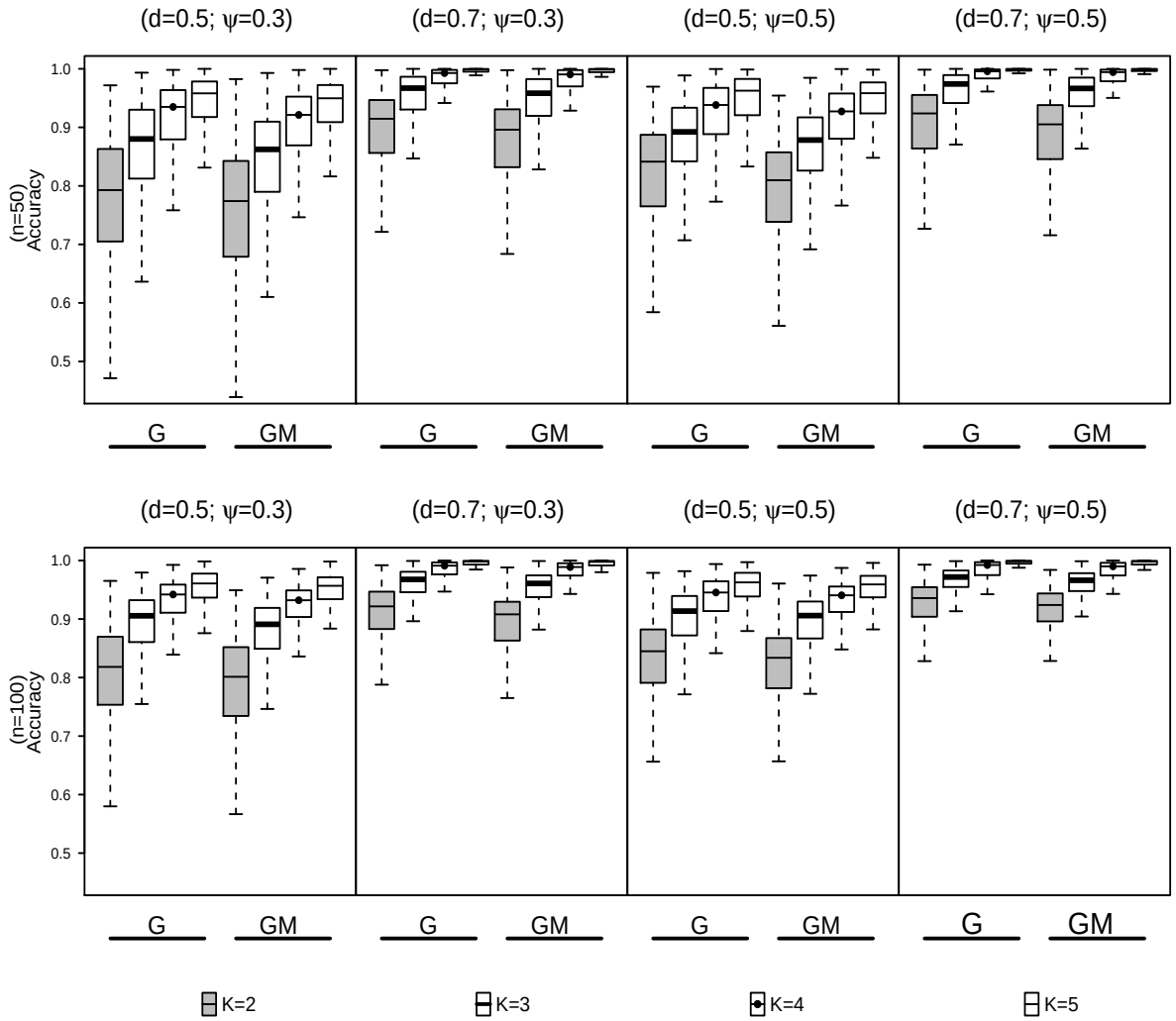

Figure 11. Boxplots of the accuracy measurements for the predictive distribution of the proportions of occupied sites associated with the Gaussian (G) and Gaussian Mixture (GM) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2, 3, 4$ and 5.

rectangular grids across South Africa. Data was extracted for three bird species from this database and included all grid-cells in Gauteng and its surroundings, in South Africa, that contained a minimum of three surveys. It was decided to restrict the maximum number of surveys to five per site. Below we report on the results for three of the species investigated namely, 1. Crowned Lapwing (Vanellus coronatus), 2. Cape Turtle-Dove (Streptopelia capicola), and 3. Southern Red Bishop (Euplectes orix).

A model was fitted with one detection covariate and one occupancy covariate, both of which were standardised to have a mean of zero and variance of one. The detection covariate used was the number of species observed by the birder (denoted as *nspp*) while the occupancy probability was modelled as a function of the ratio of potential to realised

evapotranspiration ($AETPET$).

MCMC sampling was undertaken using the R code supplied in Dorazio and Rodriguez (2012) while all variational approximations were performed using the authors' code. The MCMC estimation entails using the same prior specifications used in Section 3.2.2. One long chain of 100 000 iterations were used with a burn-in sample of 50 000.

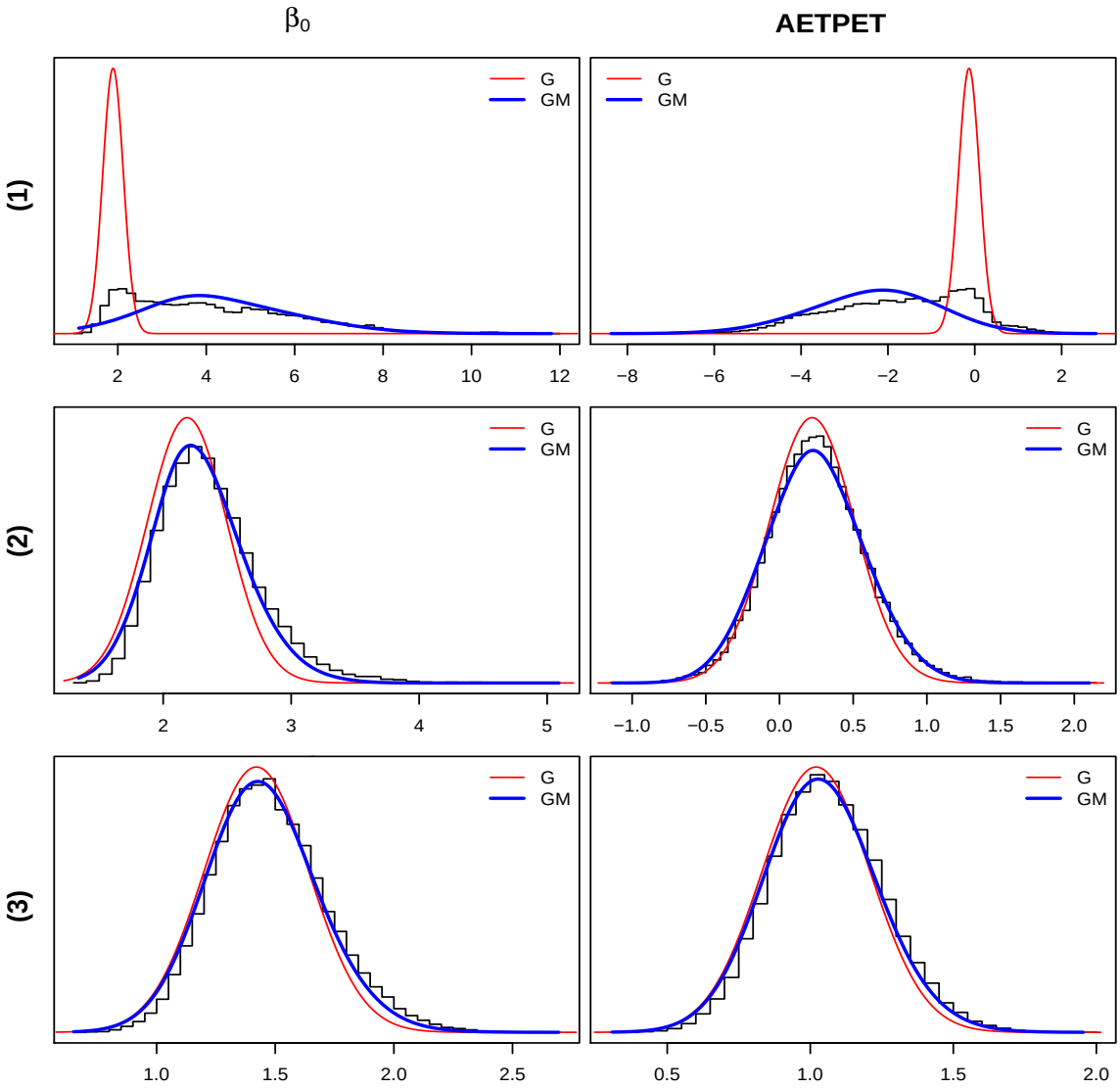

Figure 12. A comparison between the VB distributions (solid line) and the posterior distributions obtained using MCMC (the histogram) for the regression parameters of the occupancy process for the different bird species (denoted as (1) = Crowned Lapwing, (2) Cape Turtle-Dove and (3) Southern Red Bishop).

The results in Table 6 and Figures 12 and 15 indicate that when the posterior distribution of the occupancy and detection regression effects are almost symmetric (with one mode) both VB methods accurately approximate the posterior mean and standard deviations of those obtained using MCMC. The GM method is better at approximating the posterior distributions of the parameters of the SSO model since the method approximates the

Table 6. Parameter estimates of the single season occupancy models fitted using VB and MCMC.

		VB - G		**VB - GM**		**MCMC**		
Species	Covariate	Mean	Std	Mean	Std	Mean	Std	Median
Crowned Lapwing	α_0	0.731	0.068	0.697	0.067	0.702	0.070	0.701
	α_1	0.349	0.070	0.358	0.069	0.358	0.070	0.357
	β_0	**1.893**	**0.232**	**4.292**	**1.706**	**4.136**	**1.875**	3.830
	β_1	**-0.132**	**0.242**	**-2.330**	**1.499**	**-1.752**	**1.630**	-1.586
Cape Turtle Dove	α_0	1.146	0.077	1.147	0.078	1.148	0.077	1.147
	α_1	0.199	0.077	0.201	0.078	0.198	0.078	0.197
	β_0	2.187	0.311	2.279	0.363	2.363	0.395	2.315
	β_1	0.221	0.289	0.241	0.336	0.256	0.331	0.245
Southern Red Bishop	α_0	0.605	0.071	0.606	0.071	0.603	0.073	0.603
	α_1	0.459	0.078	0.460	0.079	0.461	0.079	0.460
	β_0	1.421	0.221	1.446	0.240	1.480	0.242	1.463
	β_1	1.020	0.189	1.039	0.201	1.062	0.197	1.054

Table 7. Weight contributions of the GM method.

		Weights					
Effect	Species	1	2	3	4	5	6
Detection	Crowned Lapwing	0.99	0.01	-	-	-	-
	Cape Turle-dove	0.986	0.014	-	-	-	-
	Southern Red Bishop	0.987	0.013	-	-	-	-
Occupancy	Crowned Lapwing	0.733	0.179	0.039	$< 10^{-4}$	0.044	0.005
	Cape Turle-dove	0.768	0.121	0.087	0.008	0.016	-
	Southern Red Bishop	0.939	0.058	0.002	$< 10^{-4}$	-	-

posterior distributions using a linear combination of Gaussian distributions. We however found that often the posterior distributions were not severely skewed which lead to many of the components of the linear Gaussian mixture distributions having low contributions to the overall approximate distribution (see Table 7). Case (2) and (3) in Figure 12 illustrates this characteristic. Case (1) has been included to display a scenario where the Gaussian VB approximation fails to provide a good approximation to the posterior distribution of the regression effects. This occurs since the method approximates the posterior distributions about one of its modes.

The VB approximations of the predictive distribution of the PAO are very similar for Cases (2) and (3) since the approximating posterior distributions of the regression effects are very similar. The approximations however are noticeably different for Case (1). The VB estimate of the predictive distribution are functions of the posterior moments of the approximating distributions and thus by design will have shorter tails than the true predictive distribution of the PAO. As an example of the hybrid method (proposed in Section 3.3.1) data for the Southern Red Bishop was used to obtain estimates of the

predictive distribution of the PAO. Based on the results in Figure 13 it appears as if the hybrid method is a viable alternative to both VB methods if the PAO is of interest.

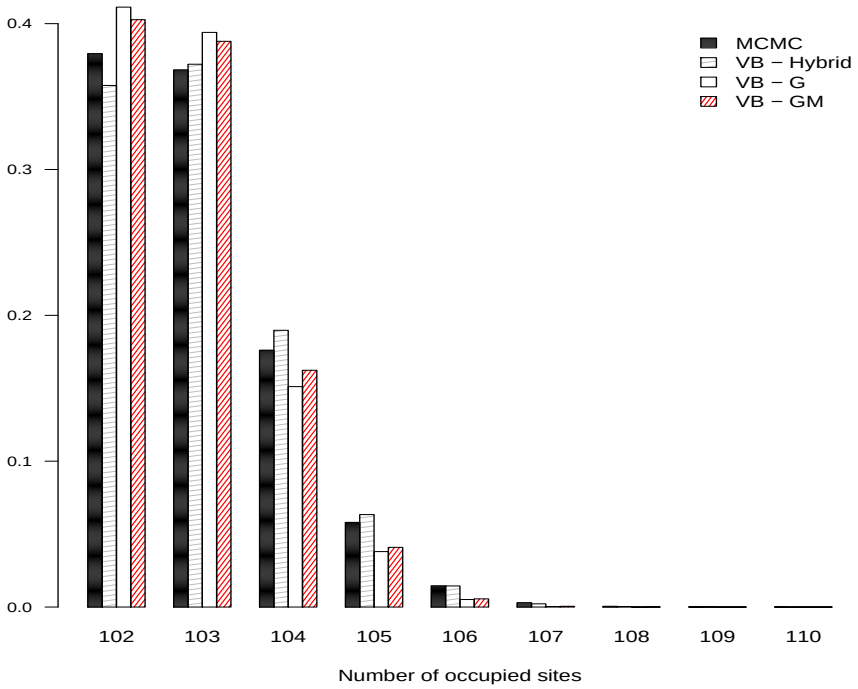

Figure 13. Predictive distribution of the proportions of occupied sites using the VB methods and the MCMC method for the Southern Red Bishop.

3.4 Conclusion

Occupancy models constitute an important class of statistical models used to examine occupancy patterns of species that are not detected perfectly. These models are often analysed using Bayesian methods. In this article two algorithms were developed that approximate the posterior distributions of the regression effects of the single season occupancy model as well as the proportion of occupied sites in a region. We found that the use of a linear mixture of Gaussian distributions could accurately approximate regression effects of the model and performed better than a single multivariate Gaussian distribution. Both methods however produced very similar approximations to the predictive posterior distribution of the proportion of occupied sites in a region.

Appendix

3.A Gaussian Mixture (GM) distribution.

Below some of the properties of the Gaussian mixture (GM) distribution (Reynolds, 2015) are discussed.

A random variable $\boldsymbol{x}$ is said to be GM distributed if $\boldsymbol{x} \sim \sum_k w_{\boldsymbol{x}}^{(k)} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{x}}^{(k)})$ where $w_{\boldsymbol{x}}^{(k)} \in [0, 1]$ with $\sum_k w_{\boldsymbol{x}}^{(k)} = 1$. The distribution of $\boldsymbol{x}$ is a linear combination of Gaussian random variables where each of the components have their own parameters $\boldsymbol{\mu}_{\boldsymbol{x}}^{(k)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}}^{(k)}$ which should be estimated from the data. The mean vector of a GM distribution (denoted as $\boldsymbol{\mu}_{\boldsymbol{x}}$) is calculated as

$$\boldsymbol{\mu}_{\boldsymbol{x}} = \sum_k w_{\boldsymbol{x}}^{(k)} \boldsymbol{\mu}_{\boldsymbol{x}}^{(k)}$$

while the covariance matrix of a GM distribution (denoted as $\boldsymbol{\Sigma}_{\boldsymbol{x}}$) is calculated as

$$\boldsymbol{\Sigma}_{\boldsymbol{x}} = \left(\sum_k w_{\boldsymbol{x}}^{(k)} \left(\boldsymbol{\Sigma}_{\boldsymbol{x}}^{(k)} + \boldsymbol{\mu}_{\boldsymbol{x}}^{(k)} \boldsymbol{\mu}_{\boldsymbol{x}}^{(k)T} \right) \right) - \boldsymbol{\mu}_{\boldsymbol{x}} \boldsymbol{\mu}_{\boldsymbol{x}}^T.$$

Let $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{a}$ where $\boldsymbol{a}$ and $\boldsymbol{D}$ is a vector and matrix of constants respectively. It can then be shown that $\boldsymbol{y}$ is GM distributed as

$$\boldsymbol{y} \sim \sum_k w_{\boldsymbol{x}}^{(k)} \mathcal{N}(\boldsymbol{D}\boldsymbol{\mu}_{\boldsymbol{x}}^{(k)} + \boldsymbol{a},\ \boldsymbol{D}\boldsymbol{\Sigma}_{\boldsymbol{x}}^{(k)} \boldsymbol{D}^T).$$

3.B Lower bound of the marginal likelihood.

Let the true occupancy states be denoted as $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)^T$ while $\mathbf{1}_n$ is an $n \times 1$ vector with all elements equal to 1. $\boldsymbol{y}$ is defined to be the $N \times 1$ vector such that its elements $[\boldsymbol{y}_i]$ are $K_i \times 1$ vectors representing the detection-nondetection data of site i. Further let $\tilde{\boldsymbol{z}} = (z_1 \mathbf{1}_{K_1}^T, \ldots, z_n \mathbf{1}_{K_n}^T)^T$ be the $N \times 1$ matrix such that $\text{diag}(\tilde{\boldsymbol{z}})$ is an $N \times N$ matrix whose main diagonal entries are the elements of $\tilde{\boldsymbol{z}}$ and zeros everywhere else.

Given the above matrices the joint posterior distribution of the regression parameters and

the true occupancy states is proportional to

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y}) \propto \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) \prod_{i=1}^{n} \boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})^{z_i} (1 - \boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta}))^{1-z_i} \times \prod_{j=1}^{K_i} (z_i \boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T \boldsymbol{\alpha}))^{y_{ij}} (1 - z_i \boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T \boldsymbol{\alpha}))^{1-y_{ij}}. \tag{3.5}$$

The natural logarithm of (3.5) is

$$\ln p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y}) \propto \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \sum_{i=1}^{n} \Big(z_i \ln(\boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})) + (1 - z_i) \ln(1 - \boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})) \Big) + \sum_{i=1}^{n} \sum_{j=1}^{K_i} \Big(y_{ij} \ln(z_i \boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T \boldsymbol{\alpha})) + (1 - y_{ij}) \ln(1 - z_i \boldsymbol{\Phi}(\boldsymbol{w}_{ij}^T \boldsymbol{\alpha})) \Big).$$

Vector evaluations of $a(x)$ and $b(x)$ results in

$$a(\boldsymbol{X}) = \begin{bmatrix} \ln(\boldsymbol{\Phi}(x_1)) \\ \vdots \\ \ln(\boldsymbol{\Phi}(x_n)) \end{bmatrix} \quad \text{and} \quad b(\boldsymbol{X}) = \begin{bmatrix} \ln(1 - \boldsymbol{\Phi}(x_1)) \\ \vdots \\ \ln(1 - \boldsymbol{\Phi}(x_n)) \end{bmatrix}.$$

Using the above matrix forms, the following expressions

$$\begin{aligned} \sum_i z_i \ln(\boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})) &= \boldsymbol{z}^T a(\boldsymbol{X}\boldsymbol{\beta}), \\ \sum_i z_i \ln(1 - \boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})) &= \boldsymbol{z}^T b(\boldsymbol{X}\boldsymbol{\beta}) \text{ and} \\ \sum_i \ln(1 - \boldsymbol{\Phi}(\boldsymbol{x}_i^T \boldsymbol{\beta})) &= \mathbf{1}_n^T b(\boldsymbol{X}\boldsymbol{\beta}) \end{aligned}$$

are obtained such that the natural logarithm of (3.5) can be re-expressed using matrix notation as

$$\ln p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y}) \propto \ln \pi(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \boldsymbol{z}^T (a(\boldsymbol{X}\boldsymbol{\beta}) - b(\boldsymbol{X}\boldsymbol{\beta})) + \mathbf{1}_n^T b(\boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{y}^T \text{diag}(\tilde{\boldsymbol{z}})(a(\boldsymbol{W}\boldsymbol{\alpha}) - b(\boldsymbol{W}\boldsymbol{\alpha})) + \mathbf{1}_N^T \text{diag}(\tilde{\boldsymbol{z}}) b(\boldsymbol{W}\boldsymbol{\alpha}).$$

Let $q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z})$ represent the approximating distribution of $p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{y})$ and assume that it can be factorized into $q(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{z}) = q(\boldsymbol{\alpha})q(\boldsymbol{\beta})q(\boldsymbol{z}) = q(\boldsymbol{\alpha})q(\boldsymbol{\beta}) \prod_i q(z_i)$.

The lower bound of the marginal log-likelihood is calculated as

$$\ln p(\boldsymbol{y}) \geq \mathbb{E}(\ln p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})) - \int q(\boldsymbol{\alpha}, \boldsymbol{\beta}) \ln q(\boldsymbol{\alpha}, \boldsymbol{\beta}) d\boldsymbol{\alpha} d\boldsymbol{\beta} - \sum_i q(z_i) \ln q(z_i) \tag{3.6}$$

$$\geq \mathbb{E}(\ln p(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta})) + H(q(\boldsymbol{\alpha})) + H(q(\boldsymbol{\beta})) - \sum_i q(z_i) \ln q(z_i) \tag{3.7}$$

where the entropy of $q(\boldsymbol{\theta})$ is calculated as $H(q(\boldsymbol{\theta})) = -\int q(\boldsymbol{\theta}) \ln q(\boldsymbol{\theta}) d\boldsymbol{\theta}$ for some general M dimensional vector $\boldsymbol{\theta}^T = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)$. The entropy calculations cannot be calculated analytically although one can generalise the results of Gershman et al. (2012) who shows that

$$\begin{aligned} H(q) &\geq -\sum_{k=1}^{K} w_{\boldsymbol{\theta}}^{(k)} \ln \left(\int q(\boldsymbol{\theta}) \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}\right) \right) d\boldsymbol{\theta} \\ &\geq -\sum_{k=1}^{K} w_{\boldsymbol{\theta}}^{(k)} \ln \left(\int \left(\sum_{j} w_{\boldsymbol{\theta}}^{(j)} \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)}\right) \right) \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}\right) d\boldsymbol{\theta} \right) \\ &\geq -\sum_{k=1}^{K} w_{\boldsymbol{\theta}}^{(k)} \ln \left(c_{\boldsymbol{\theta}}^{(k)} \right) \end{aligned}$$

where

$$\begin{aligned} c_{\boldsymbol{\theta}}^{(k)} &= \int \left(\sum_{j} w_{\boldsymbol{\theta}}^{(j)} \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)}\right) \right) \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}\right) d\boldsymbol{\theta} \\ &= \sum_{j} w_{\boldsymbol{\theta}}^{(j)} \int \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(j)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)}\right) \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}\right) d\boldsymbol{\theta} \\ &= \sum_{j} w_{\boldsymbol{\theta}}^{(j)} (2\pi)^{-M/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}|^{-0.5} \times \\ &\quad \exp\left(-\frac{1}{2} \left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(j)}\right)^T \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}\right)^{-1} \left(\boldsymbol{\mu}_{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\mu}_{\boldsymbol{\theta}}^{(j)}\right) \right). \end{aligned}$$

Note (3.3) requires the calculation of the inverse of $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(j)} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{(k)}$ when undertaking the VB iterative scheme. This calculation could become prohibitively costly and increase computing time if many covariates are included in the detection and occupancy processes. In most cases the researcher will however include a small number of explanatory variables so that the additional computation time required to undertake the inverse calculations will be minimal.

3.C Box plots of accuracy statistics.

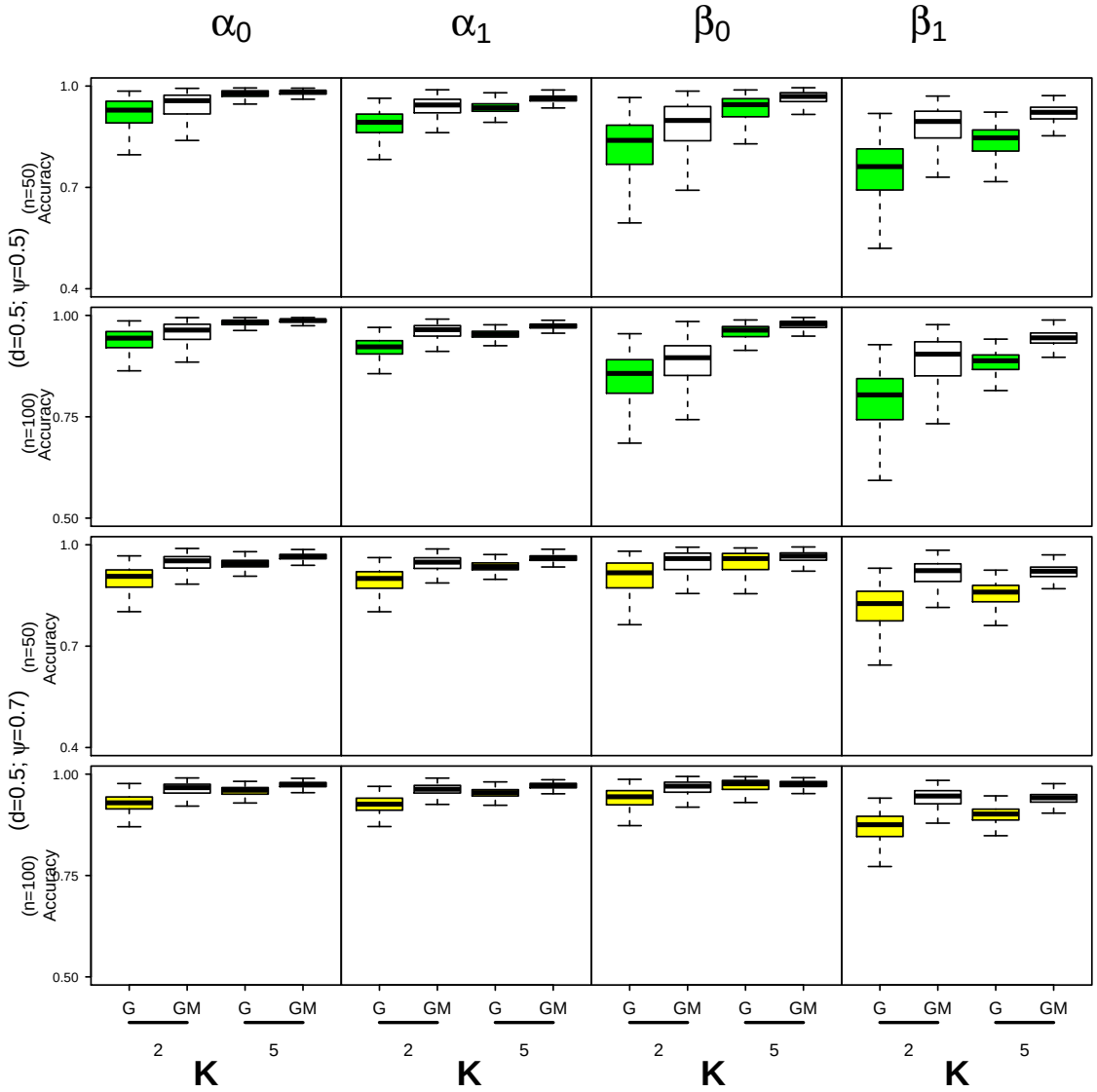

Figure 14. Box plots of the accuracy measurements for the model parameters associated with the Gaussian (G) and Gaussian Mixture (GM) based method for number of sites $n = 50,\ 100$ and number of visits to each site $K = 2,\ 5$. Two detection and occupancy probability scenarios are considered here namely $(d = 0.5,\ \psi = 0.5)$ and $(d = 0.7,\ \psi = 0.5)$.

3.D Posterior distribution and VB approximations of the detection regression effects.

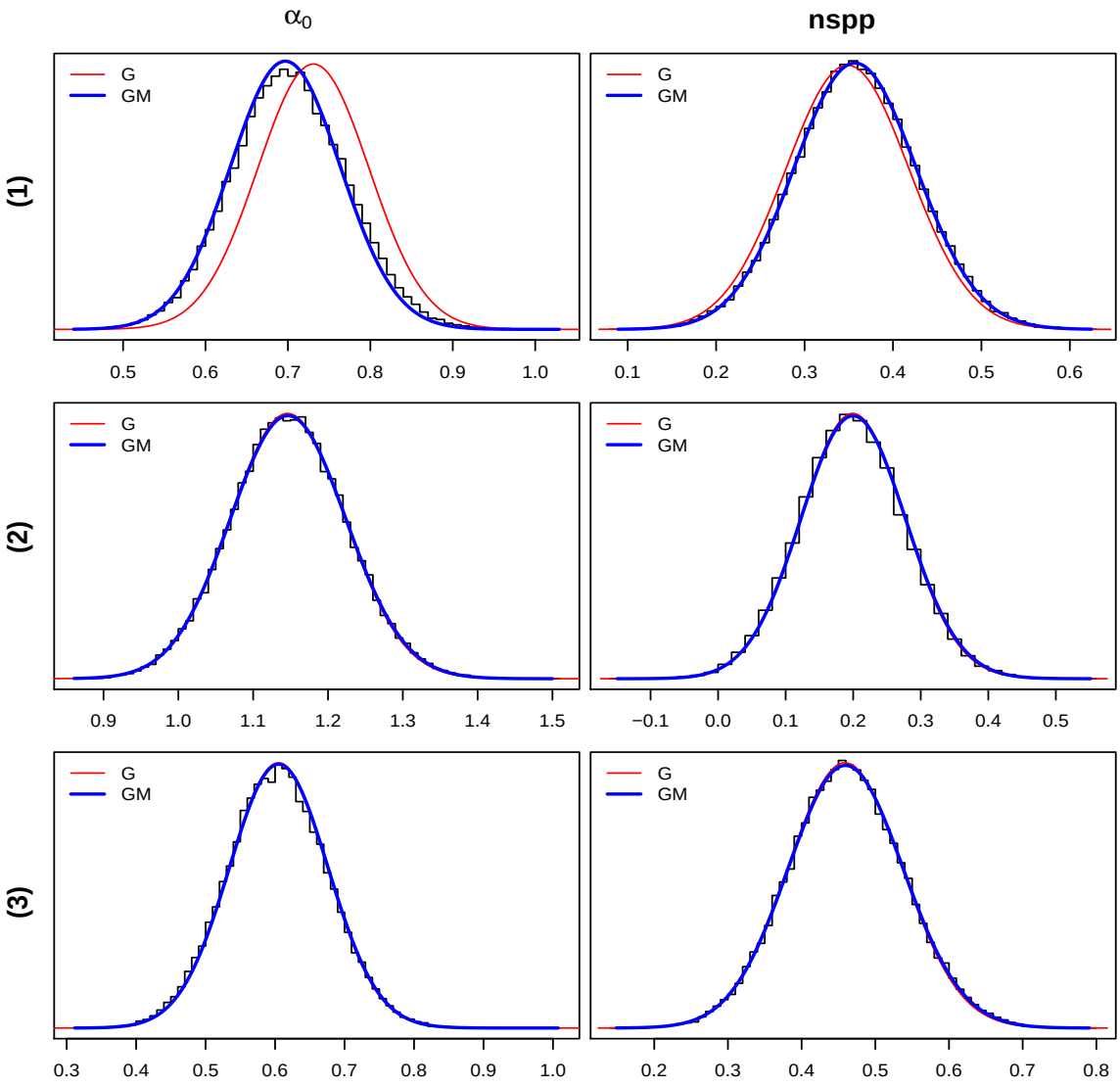

Figure 15. A comparison between the VB distributions (solid red and blue lines) and the posterior distributions obtained using MCMC (the histogram, black line) for the regression parameters of the detection process for the different bird species (denoted as (1) = Crowned Lapwing, (2) Cape Turtle-Dove and (3) Southern Red Bishop).

Appendix

Additional material for Chapter 3

The R code used to undertake the analysis can be found at `https://github.com/AllanClark/PhdAnalysisCode/tree/main/Chapter3_VBOccupancyProbit`.

Chapter 4

Efficient Bayesian analysis of occupancy models with logit link functions.

The following paper has been accepted for publication in Ecology and Evolution. Minor changes have been made to this version of the published manuscript. The complete citation for the manuscript is: Clark, A.E. and Altwegg, R., 2019. Efficient Bayesian analysis of occupancy models with logit link functions. Ecology and Evolution, 9(2), pp.756-768. Supplementary material, data files and some example code can be found at `https://datadryad.org/stash/dataset/doi:10.5061/dryad.jt2002k` while the *Rcppocc* package can be found at `https://github.com/AllanClark/Rcppocc`.

Portions of the work in this chapter was presented at the following conferences: (i) a poster presentations at *The Euring 2017 Analytical Meeting*, Barcelona, Spain, 2-7 July 2017 and *ISBA World Meeting*, Edinburgh, Scotland, 24-29 July 2018(ii) oral presentations at the *International Statistical Ecology conference*, St Andrews, Scotland, 2-6 July 2018 and the *South African Statistical Association conference*, Johannesburg, South Africa, 26-29 November 2018.

The R code used to undertake the analysis and simulation study can be found at `https://github.com/AllanClark/PhdAnalysisCode`.

Take note that the references for the article are all displayed at the end of the thesis.

'Allan Ernest Clark' wrote the paper and performed the mathematical development and analysis. 'Res Altwegg' and 'Allan Ernest Clark' reviewed the manuscript extensively. Both authors have read and approved the manuscript.
Some corrections to the manuscript in this version:

- The reference 'Balakrishnan (2013)' is corrected to 'Balakrishnan (1991)'.
- V_i has been changed to K_i for consistency throughout the thesis.
- Bern has been changed to Bernoulli.
- Vectors and matrices are defined using square braces instead of round braces.
- The terms 'detection/non-detection' and 'logit' has been changed to 'detection-nondetection' and 'logistic' for consistency throughout the thesis.

Efficient Bayesian analysis of occupancy models with logit link functions.

Allan E. Clark [1,2], Res Altwegg[1,2],

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.

Abstract

Occupancy models (MacKenzie et al., 2002) were developed to infer the probability that a species under investigation occupies a site. Bayesian analysis of these models can be undertaken using statistical packages such as *WinBUGS*, *OpenBUGS*, *JAGS* and more recently *Stan*, however, since these packages were not developed specifically to fit occupancy models, one often experiences long run-times when undertaking an analysis. Bayesian spatial single season occupancy models can also be fit using the R package *stocc*. The approach assumes that the detection and occupancy regression effects are modelled using probit link functions.

The use of the logistic link function however is algebraically more tractable and allows one to easily interpret the coefficient effects of an estimated model by using odds ratios, which is not easily done for a probit link function for models that do not include spatial random effects. We develop a Gibbs sampler to obtain posterior samples from the posterior distribution of the parameters of various occupancy models (nonspatial and spatial) when logit link functions are used to model the regression effects of the detection and occupancy processes.

We apply our methods to data extracted from the 2$^{\text{nd}}$ Southern African Bird Atlas Project to produce a species distribution map of the Cape weaver (*Ploceus capensis*) and helmeted guineafowl (*Numida meleagris*) for South Africa. We found that the Gibbs sampling algorithm developed produces posterior samples that are identical to those obtained when using *JAGS* and *Stan* and that in certain cases the posterior chains mixes much faster than those obtained when using *JAGS*, *stocc* and *Stan*.

Our algorithms are implemented in the R package, *Rcppocc*. The software is freely available and stored on GitHub (`https://github.com/AllanClark/Rcppocc`).

Key words

Bayesian spatial occupancy model, occupancy model, restricted spatial regression, imperfect detection, Rcppocc.

4.1 Introduction

Occupancy models are an important statistical technique that was developed to make use of detection-nondetection data to infer the probability that a species under investigation occupies a site. When an occupancy study is undertaken, n_s sites are visited a number of times to estimate the occupancy probability ($\boldsymbol{\psi}$) and conditional detection probability ($\boldsymbol{p}$) of a species associated with each site in a region. The method can be viewed as an extension of logistic regression and allows one to estimate the occupancy probability at sites where none of the species being investigated have been detected. The model is formulated hierarchically, using Bernoulli random variables to specify the occupancy and detection processes respectively which can be modelled using site-specific and survey-specific explanatory variables respectively (MacKenzie et al., 2002). Johnson et al. (2013) note that occupancy models produce 'unbiased inference when occupancy observations at nearby units are conditionally independent given any available covariates' but stress that 'spatial autocorrelation may lead to biases and overestimated precision' of regression effects. This observation has lead to the development of various models to account for spatial autocorrelation in ecological survey data (Hoeting et al., 2000; Hooten et al., 2003; Gardner et al., 2010; Aing et al., 2011) and have extensively been used to guide environmental monitoring and assessment programs globally.

A number of methods have been used to fit occupancy models to data. These include maximum likelihood (MacKenzie et al., 2002); penalized maximum likelihood (Moreno and Lele, 2010; Hutchinson et al., 2015), Bayesian methods that employ *WinBUGS*, *OpenBUGS*, *JAGS* or *Stan* as well as approximate methods such as those developed by Clark et al. (2016). Recently Dorazio and Rodriguez (2012) and Johnson et al. (2013) developed Gibbs algorithms to obtain posterior samples for the parameters of a nonspatial and spatial single season occupancy (SSO) model respectively. Both approaches assume that detection and occupancy processes are modelled using probit link functions, which enables the use of data augmentation (Tanner and Wong, 1987) to obtain closed form expressions of the conditional posterior distributions of the parameters of the occupancy model.

Given that the probit and logistic functions are very similar and only differ in respect of the tails of the functions, analysis undertaken using either of the functions should produce similar occupancy and conditional detection probabilities (Dorazio and Rodriguez, 2012). However, the use of the logistic link function is algebraically more tractable and allows one to easily interpret the coefficient effects of an estimated model by using odds ratios, which is not easily done for a probit link function. This observation is particularly true for the nonspatial SSO model since no spatial random effects are included in this model; however, when spatial random effects are included in the model, the interpretation of the regression effects can be difficult (Boehm et al., 2013).

The paper commences with a brief discussion of the link between logistic regression and occupancy models. Thereafter, we discuss the formulation of various popular Bayesian spatial occupancy models and develop a Gibbs sampling algorithm for a particular spatial occupancy model when the regression effects of the occupancy and detection processes are modelled using logistic link functions. Before concluding, we analyse two detection-nondetection data sets of South African bird species to illustrate the methods developed in the paper. An R package (*Rcppocc)* has been developed to fit SSO models using Gibbs sampling which can be obtained at `https://github.com/AllanClark/Rcppocc`.

4.2 Material and Methods

4.2.1 Logistic regression and occupancy models

Assume that n_s sites are surveyed a number of times and detection-nondetection data are collected at all sites. Denote the observed data as a ragged matrix $\boldsymbol{y} = [y_{ij}]$ where $y_{ij} = 1$ if the species under investigation has been observed at site i during survey j and $y_{ij} = 0$ otherwise. Let the vector $\boldsymbol{z}$ represent the true species occupancy at the sites considered such that $z_i = 1$ if the species occupies site i and $z_i = 0$ if it does not occupy site i. The SSO model can be represented using the following hierarchical model, $z_i|\psi_i \sim \text{Bernoulli}(\psi_i)$, $y_{ij}|z_i, p_{ij} \sim \text{Bernoulli}(z_i p_{ij})$ for all sites $i = 1, \ldots, n_s$; for all surveys $j = 1, \ldots, K_i$ (Royle and Dorazio, 2008). The variable ψ_i denotes the occupancy probability at site i while $p_{ij} = \Pr(y_{ij} = 1|z_i = 1)$ denotes the conditional probability of detecting the species during the j^{th} survey of site i given that the species is present at site i. In what follows we assume that the conditional detection and occupancy regression effects ($\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$) are modelled using logit link functions such that $\text{logit}(\psi_i) = \boldsymbol{x}_i^T\boldsymbol{\beta}$ and $\text{logit}(p_{ij}) = \boldsymbol{w}_{ij}^T\boldsymbol{\alpha}$, where $\boldsymbol{x}_i^T$ and $\boldsymbol{w}_{ij}^T$ are row-vectors in design matrices, $\boldsymbol{X}$ (occupancy) and $\boldsymbol{W}$ (detection) respectively (as defined in Clark et al. (2016)).

The joint posterior distribution of the parameters of the model is

$$[\boldsymbol{z}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\boldsymbol{y}] \propto \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta}) \left(\prod_{i=1}^{n_s} \psi_i^{z_i}(1-\psi_i)^{1-z_i} \right) \prod_{\{i:z_i=1\}} \prod_j p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}$$

where $\pi(\boldsymbol{\alpha})$ and $\pi(\boldsymbol{\beta})$ are the prior distributions of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ respectively. A directed acyclic graph of the above problem is displayed in Figure 16 below.

A Gibbs sampling algorithm for the parameters of this model requires sampling from $[\boldsymbol{\beta}|\boldsymbol{z}]$, $[\boldsymbol{\alpha}|\boldsymbol{z}, \boldsymbol{y}]$, and $[z_i = 1|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{y}]$ for all sites where the species has not been observed. The

first two conditional distributions have the following form,

$$[\boldsymbol{\beta}|\boldsymbol{z}] \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n_s} \psi_i^{z_i}(1-\psi_i)^{1-z_i} \text{ and} \tag{4.1}$$

$$[\boldsymbol{\alpha}|\boldsymbol{z},\boldsymbol{y}] \propto \pi(\boldsymbol{\alpha}) \prod_{\substack{i \\ \{i:z_i=1\}}} \prod_{j} p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}. \tag{4.2}$$

Notice that equations (4.1) and (4.2) are of the same form as the posterior distributions of the regression effects of a logistic regression model and therefore we adapt a Gibbs sampling scheme for logistic regression models to address the problem of obtaining posterior samples for the parameters of an occupancy model.

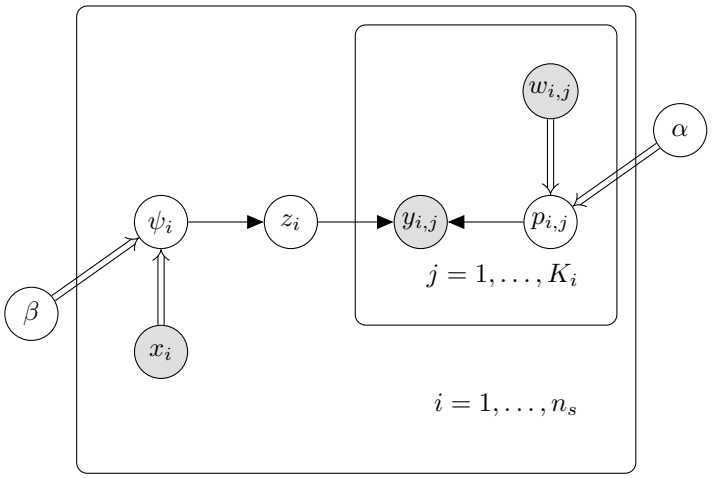

Figure 16. A directed acyclic graph illustrating the dependencies between the parameters and observed data for an SSO model. Shaded nodes represents observed data while all latent parameters are represented using unshaded nodes. Deterministic relationships are represented using double arrows while all stochastic relationships are represented using a single arrow.

In a logistic regression context, Polson et al. (2013) show that posterior samples of the regression effects can be obtained by sampling from the conditional distributions of Pólya-Gamma random variables and multivariate Gaussian distributions in turn. Their method is similar to that of Albert and Chib (1993) who developed a Gibbs algorithm to undertake probit regression, the only difference being that the sampling from truncated Gaussian distributions are replaced by sampling from Pólya-Gamma distributions. The sampling methods developed by Polson et al. (2013) is exact since their Pólya-Gamma sampling method is uniformly ergodic and converges to the correct posterior distribution (Choi and Hobert, 2013).

For the SSO model, the conditional posterior distributions of $\boldsymbol{\beta}|\boldsymbol{\omega}_{\boldsymbol{\beta}}, \boldsymbol{y}$ are derived by introducing Pólya-Gamma latent variables, $\boldsymbol{\omega}_{\boldsymbol{\beta}}$, and noting that the contribution of the i^{th}

observation to a Bernoulli likelihood can be re-expressed as

$$\psi_i^{z_i}(1-\psi_i)^{1-z_i} = \frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)^{z_i}}{1+\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}\right)} = \exp(\kappa_i \boldsymbol{x}_i^T\boldsymbol{\beta}) \int \exp\left(-\frac{\omega_{i,\boldsymbol{\beta}}}{2}(\boldsymbol{x}_i^T\boldsymbol{\beta})^2\right) p(\omega_{i,\boldsymbol{\beta}}|1,0) d\omega_{i,\boldsymbol{\beta}},$$

where $\kappa_i = z_i - 0.5$ and $p(\omega_{i,\boldsymbol{\beta}}|1,0)$ is the probability density function of a Pólya-Gamma distribution with parameters 1 and 0 (Polson et al., 2013). The conditional posterior distribution of $\boldsymbol{\alpha}$ is derived by using the same manipulation of the Bernoulli likelihood.

In Appendix 4.A we discuss the existing Gibbs algorithms used for undertaking logistic regression and demonstrate the use of the Pólya-Gamma (PG) method by developing two Gibbs sampling algorithms for the parameters of SSO models. In Table 8, we summarise the Gibbs algorithms for an SSO model when using the PG method but leave the details regarding the algorithm in Appendix 4.B and Appendix 4.C. We use the notation '$a \sim \text{PG}(b,c)$' to indicate that the random variable a is a Pólya-Gamma random variable with parameters b and c. Take note that the algorithm is identical to that developed by Dorazio and Rodriguez (2012) except that the sampling from truncated Gaussian distributions are replaced by sampling from Pólya-Gamma distributions.

Table 8. The Gibbs algorithm for undertaking a SSO model using the 'PG' method. (See Appendix 4.C for the details pertaining to the parameter matrices of the conditional posterior distributions.)

1: Set starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{z}$.
2: **for** (iterations $= 1, \ldots,$ simulation runs) **do**
3: **for** $(i = 1, \ldots, n_s)$ **do**
4: - Generate $\omega_{i,\boldsymbol{\beta}} \sim \text{PG}(1, \boldsymbol{x}_i^T\boldsymbol{\beta})$.
5: **end for**
6: - Generate $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$.
7: **for** $((ij) = \{z_i = 1\})$ **do**
8: - Generate $\omega_{ij,\boldsymbol{\alpha}} \sim \text{PG}(1, \boldsymbol{w}_{ij}^T\boldsymbol{\alpha})$.
9: **end for**
10: - Generate $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$.
11: - Generate $z_i \sim \text{Bernoulli}(\tilde{\psi}_i)$.
12: **end for**

4.2.2 Bayesian spatial SSO models.

Spatial generalized linear mixed models (SGLMM) are an extension of the general linear model (Nelder and Wedderburn, 1972) that allow the link function of the expected value of the random variable under investigation to be modelled as a function of a spatial random variable/s. The formulation was first developed by Besag et al. (1991) and has

been extensively used in areas such as agriculture (Besag and Higdon, 1999), biostatistics (Gelfand and Vounatsou, 2003; Waller and Gotway, 2004), ecology (Lichstein et al., 2002) and species distribution modelling (Latimer et al., 2006; Hooten et al., 2003; Gelfand et al., 2005; Drouilly et al., 2018).

The paper by Gelfand et al. (2005) lead to the development of the R package *hSDM* (Vieilledent et al., 2014) in which the *hSDM.siteocc.iCAR* function can be used to fit a particular spatial occupancy model to detection-nondetection data. A region under investigation is sub-divided into n_s grid-cells and are surveyed a number of occasions. The model is formulated using Bernoulli latent random variables $\boldsymbol{z} = [z_1, \ldots, z_{n_s}]^T$. Formally, $z_i|\psi_i \sim \text{Bernoulli}(\psi_i)$ with $\text{logit}(\psi_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \rho_i,$ for all $i = 1, \ldots, n$, where $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_{n_s}]^T$ is a multivariate Gaussian random vector with mean $\mathbf{0}$ and correlation matrix defined using the neighbourhood structure of the grid-cells. The observation process is specified as in the nonspatial model. The documentation of the function indicates that posterior samples of the parameters of the model are obtained using the C programming language and utilizes an adaptive Metropolis algorithm (Metropolis et al., 1953; Robert and Casella, 1999).

Johnson et al. (2013) develop two spatial occupancy models. They assume that probit link functions are used to model both the occupancy and detection processes and thereby rely on data augmentation to develop a Gibbs sampling algorithm to sample from the posterior distribution of the parameter of the models. For the probit case, the occupancy probability of a particular grid-cell (for the standard occupancy model) is calculated as $\Phi(\boldsymbol{x}_i^T\boldsymbol{\beta}) = \Pr(z_i = 1)$. In a Bayesian context, such a probit model is formulated by defining a latent Gaussian random variable, $\tilde{z}_i$ with mean 0 and variance equal to 1 such that $\Pr(z_i = 1) = \Pr(\tilde{z}_s > 0)$. In the first of their models they allow $\tilde{z}_s$ to be spatially correlated such that $\tilde{z}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \eta_i + \epsilon_i$. $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_{n_s}]^T$ is defined as $\boldsymbol{\rho}$ above while $\epsilon_i \sim \mathcal{N}(0, 1),$ for all $i = 1, \ldots, n_s$.

Often the spatial random effects and fixed-effects of a model are collinear when spatially varying covariates are included as fixed-effects (Reich et al., 2006; Hodges and Reich, 2010; Hughes and Haran, 2013; Hanks et al., 2015). The suggested solution to this problem was to include spatial random effects in the model specification that are orthogonal to the fixed effects and is known as *restricted spatial regression* (RSR). The second spatial model developed by Johnson et al. (2013) uses this method and redefines $\tilde{z}_i$ as $\tilde{z}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{k}_i^T\boldsymbol{\theta} + \epsilon_i$, where $\boldsymbol{k}_i^T$ is a row-vector of the design matrix $\boldsymbol{K}$. The spatial random effects are modelled as

$$\begin{aligned}\boldsymbol{\theta}|\tau &\sim \mathcal{N}\left(\mathbf{0}_r,\ \frac{1}{\tau}\left(\boldsymbol{K}^T\boldsymbol{Q}\boldsymbol{K}\right)^{-1}\right) = \mathcal{N}\left(\mathbf{0}_r,\ \frac{1}{\tau}\boldsymbol{M}\right),\\ \tau &\sim \mathcal{G}(i_1,\ i_2) \text{ and } \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}_n,\ \boldsymbol{I}_n).\end{aligned}$$

$\boldsymbol{Q}$ is a $n \times n$ ICAR precision matrix (Besag and Kooperberg, 1995) obtained using surveyed and unsurveyed locations, τ is a spatial precision parameter and i_1 and i_2 are known constants. Kelsall and Wakefield (1999) have suggested setting these parameters to 0.5 and 0.005 respectively such that the prior mean of τ is 1000. The matrix $\boldsymbol{K}$ consists of the first r ($r << n$) eigenvectors of $\boldsymbol{\Omega} = n\boldsymbol{RAR}/\mathbf{1}^T\boldsymbol{A}\mathbf{1}$ where $\boldsymbol{R} = \boldsymbol{I}_n - \boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T$ and $\boldsymbol{A}$ is an association matrix with $(ij)^{\text{th}}$ entry $A_{ij} = 1$ if sites i and j are neighbours and zero otherwise.

In our formulation of the spatial occupancy model we model the occupancy probabilities at all grid-cells as

$$\text{logit}(\psi_i) = \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{k}_i^T\boldsymbol{\theta}, \tag{4.3}$$

where $\boldsymbol{K}$ and $\boldsymbol{\theta}$ are defined above. We make use of Pólya-Gamma random variables to obtain the conditional distributions of the parameters of the above spatial occupancy model. The conditional distributions are very similar to those obtained for the SSO model although here we require the conditional posterior distribution of additional parameters ($\boldsymbol{\theta}$ and τ). A directed acyclic graph of the spatial SSO model is displayed in Figure 17 below while in Table 9, we summarise the Gibbs algorithms for a spatial SSO model which employs equation (4.3) when using the PG method. The details regarding the algorithm can be found in Appendix 4.D.

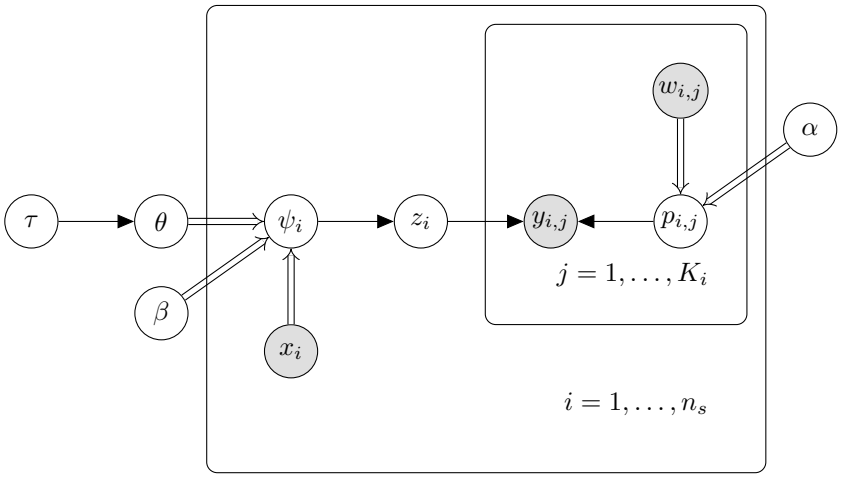

Figure 17. A directed acyclic graph illustrating the dependencies between the parameters and observed data for a spatial SSO model. Shaded nodes represents observed data while all latent parameters are represented using unshaded nodes. Deterministic relationships are represented using double arrows while all stochastic relationships are represented using a single arrow.

Table 9. The Gibbs algorithm for undertaking a spatial SSO model. (See Appendix 4.D for the details pertaining to the parameter matrices of some of the conditional posterior distributions.)

1: Set starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, τ, $\boldsymbol{\theta}$ and $\boldsymbol{z}$.

2: **for** (iterations $= 1, \ldots,$ simulation runs) **do**

3: **for** $(i = 1, \ldots, n_s)$ **do**

4: - Generate $\omega_{i,\boldsymbol{\beta}} \sim \text{PG}(1, \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{k}_i^T\boldsymbol{\theta})$.

5: **end for**

6: - Generate $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta,\ \boldsymbol{\Sigma}_\beta)$.

7: - Generate $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$.

8: - Generate $\tau \sim \mathcal{G}\left(\frac{r}{2} + i_1,\ \frac{\boldsymbol{\theta}^T\boldsymbol{M}^{-1}\boldsymbol{\theta}}{2} + i_2\right)$.

9: **for** $((ij) = \{z_i = 1\})$ **do**

10: - Generate $\omega_{ij,\boldsymbol{\alpha}} \sim \text{PG}(1, \boldsymbol{w}_{ij}^T\boldsymbol{\alpha})$.

11: **end for**

12: - Generate $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$.

13: - Generate $z_i \sim \text{Bernoulli}(\tilde{\psi}_i)$.

14: **end for**

4.2.3 Applications

To demonstrate our methods, we used detection-nondetection data extracted from the 2nd Southern African Bird Atlas Project (SABAP2) database to produce a species distribution map of the Cape weaver (*Ploceus capensis*) and helmeted guineafowl (*Numida meleagris*) for South Africa. SABAP2 divides southern Africa into a continuous grid of $5' \times 5'$ and relies on citizen scientists to collect checklists of bird species for each grid cell. Birders are requested to spend at least two hours on each checklist in which they undertake *intense birding* and record all species they observe and the order in which they are observed. For this analysis, we aggregated the data to quarter-degree grid cells. We used data that span South Africa and contained a minimum of three and a maximum of fifty surveys during 2016 (January-December) in the analysis. Covariate information at unsurveyed locations was included in the analyses to obtain occupancy estimates that span South Africa. All covariates were centred and standardised.

In our analysis, we fitted a nonspatial and spatial SSO model with one detection covariate and two occupancy covariates. The detection covariate used was the number of species observed by the birder (denoted as *nspp*) while the occupancy covariates were functions of seven climate variables. It is assumed that the more species the birder observes while birding, the more more likely they are to observe the particular species being analysed such

that a positive detection regression effect is expected. The aim of the analysis is not to obtain *the best* occupancy model for the particular data sets but rather to highlight the use of the developed Gibbs sampling algorithm for fitting RSR occupancy models to the data using different software programs and different sampling methods. We specifically consider the Gibbs sampling algorithm by Johnson et al. (2013) (probit link functions), our Gibbs algorithm (logit link functions), *JAGS* as well as *Stan* (which uses a *no-U-turn* Hamiltonian Monte Carlo sampler (Hoffman and Gelman, 2014) to sample from the parameters of a posterior distribution).

The climate variables (see Figure 23 in Appendix 4.F) all form part of a data set used by Huntley et al. (2006) to model bird distributions in Southern Africa. The variables included two measures of annual temperature that related to thermal sums above 0 and 5 degrees centigrade; two measures related to the mean temperature of the coldest and warmest month respectively; the ratio of potential to realized evapotranspiration as well as two measures that relates to the intensity of the dry and wet season respectively. The climate variables are highly correlated with two of the variables having variance inflation factors in excess of 3 000 (see Table 13 and Table 14 in Appendix 4.E). Because of this fact, it was decided to extract two principal components from the design matrix that consisted of the centred and standardised climate variables. These principal components explain 90% of the variation in the design matrix (see Table 15 in Appendix 4.E) and can tentatively be interpreted as a temperature related factor and a climate intensity factor respectively.

We follow Hughes and Haran (2013) and retain 10% of the eigenvalues (λ_i $i = 1, \ldots, n$) of $\boldsymbol{\Omega}$. In a similar context, Johnson et al. (2013) suggests selecting a RSR model with $\lambda_i \geq 0.5$ which suggests including at most 237 eigenvectors into the spatial portion of the model. Experimentation with different values of r between 150 and 230 demonstrated no significant difference to the results we report here.

The following prior distributions were used for the parameters of the spatial SSO model: $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0}, 1000\boldsymbol{I}_2)$, $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 1000\boldsymbol{I}_3)$ and $\tau \sim \mathcal{G}(0.5, 0.005)$. The prior specification for τ places more weight on large values of τ indicating that very little prior weight is placed on the spatial random effects of the model. Broms (2013) performed a simulation study and found that the RSR model results are not sensitive to the prior specification of τ and thus we have not done any analysis to test the sensitivity of our results to the prior specification of τ. All MCMC sampling was undertaken using the R packages, *stocc*, *jagsUI* (Kellner, 2014) in combination with *JAGS* 4.2.0 (Plummer, 2003), *rstan* in combination with *Stan* 2.17.3 as well as the authors' code[1]. All calculations were performed on a Windows 10 Pro

[1] An R package has been developed to fit these models using MCMC. All code can be obtained from `https://github.com/AllanClark/Rcppocc`. Appendix 4.I includes a worked example explaining how to run RSR models using *stocc*, *Stan* and *Rcppocc*.

desktop computer which had an Intel(R) Core(TM) i7-6900 processor with 64 GB of RAM. One chain of 70 000 iterations was run. The first 20 000 samples were discarded as burn-in samples while the remaining samples were retained. Experimentation and an examination of the Geweke convergence diagnostic statistics (Geweke, 1992) and trace plots obtained by running three parallel chains using *Rcppocc* displayed that the MCMC chains converged using these numbers of iterations. The posterior samples were not thinned (Link and Eaton, 2012).

4.3 Results

From the analysis of both data sets we observe that the Gibbs algorithm developed for the spatial occupancy model produces identical posterior distributions to those obtained when using *JAGS* and *Stan* (see Figures 24 and 25 in Appendix 4.G). In both data sets the posterior samples of the detection regression effects exhibit good mixing where the lagged sample autocorrelations of the posterior samples approaches zero within 5 lags. The posterior samples of the occupancy regression effects as well as the precision of the spatial random effect (τ) exhibits slower mixing when using *stocc*, *JAGS* and *Rcppocc* (denoted as 'PG' in Figures 26 and 27 in Appendix 4.H) while *Stan* produced a posterior chain that mixed well. We observe that *stocc* produced posterior samples for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and τ that had the largest levels of autocorrelation amongst all of the methods considered (when fitting the RSR model).

Table 10. Posterior run-times for the Bayesian spatial occupancy models as well as the ESR (per minute) for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and τ.

Species	Method	Time (min)	α_0	α_1	β_0	β_1	β_2	τ
Cape weaver	*stocc* (RSR)	27.00	552.71	725.59	12.03	7.83	22.37	11.35
	stocc (ICAR)	136.76	104.41	142.57	0.25	0.10	0.15	0.14
	JAGS	243.55	102.62	131.42	2.69	1.71	5.09	1.97
	Stan	187.06	275.75	331.44	53.48	34.42	83.74	19.35
	Rcppocc	19.88	1682.15	1804.42	85.32	53.40	141.82	116.19
Helmeted Guinea fowl	*stocc* (RSR)	27.08	619.72	563.24	11.63	43.61	48.34	15.61
	stocc (ICAR)	165.23	65.92	97.15	0.04	0.16	0.09	0.05
	JAGS	254.49	97.86	125.35	3.08	10.84	12.06	2.66
	Stan	150.55	595.77	617.84	49.21	186.23	286.92	24.44
	Rcppocc	19.9	1888.11	1493.07	86.16	314.04	364.47	106.06

Table 10 tabulates the run-times (in minutes) and effective sampling rate (ESR[2] = the effective sample size per unit run-time) of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and τ for each of the sampling algorithms

[2]ESR = the effective sample size per unit run-time. The effective sample size for the i^{th} parameter in the model is defined as $ESS_i = \frac{M}{1 + 2\sum_{j=1}^{k} \rho_i(j)}$, where M is the number of post-burn-in samples, and $\rho_i(j)$ is the j^{th} lagged autocorrelation of parameter i (Holmes and Held, 2006). We use the *coda* package (Plummer et al., 2006) to estimate ESS_i.

used to analyse the two data sets. For completeness sake we also include the statistics related to the ICAR model when using *stocc*. We observe that *stocc* and *Rcppocc* had faster run-times than *JAGS* and *Stan*. *Rcppocc* had the fastest running-times and completed the 70 000 MCMC iterations approximately 12 times faster than *JAGS* and between $7-10$ times faster than *Stan*. *Rcppocc* has the largest ESR of all of the algorithms considered and produced ESR values which ranged between $1.5-6$ times larger than those obtained by *Stan*; $3-11$ times larger than those obtained by *stocc* and $11-60$ times larger than those obtained by *JAGS*. The ICAR models took approximately 8 times longer to run than the RSR model when using *stocc* and resulted in significantly larger levels of autocorrelation within the $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and τ chains.

Table 11. Posterior summaries of the parameters of the Bayesian nonspatial and spatial occupancy models (posterior mean, Monte Carlo standard error, Standard deviation, 2.5% and 97.5% quantiles).

Type	Species	Parameter	Mean	MCSE	Std	2.5%	97.5%
Nonspatial	Cape weaver	α_0	-0.32	0.0002	0.03	-0.38	-0.26
		α_1	0.56	0.0002	0.03	0.49	0.62
		β_0	-0.49	0.0006	0.10	-0.68	-0.30
		β_1	-0.71	0.0005	0.06	-0.84	-0.59
		β_2	-0.24	0.0004	0.06	-0.36	-0.12
	Helmeted Guineafowl	α_0	-0.10	0.0001	0.02	-0.15	-0.06
		α_1	0.78	0.0002	0.03	0.72	0.84
		β_0	-0.35	0.0006	0.06	-0.48	-0.22
		β_1	-0.36	0.0005	0.09	-0.54	-0.20
		β_2	-0.39	0.0008	0.08	-0.56	-0.24
Spatial	Cape weaver	α_0	-0.33	0.0002	0.03	-0.39	-0.27
		α_1	0.58	0.0002	0.03	0.52	0.64
		β_0	-1.54	0.0074	0.30	-2.19	-1.00
		β_1	-1.52	0.0066	0.21	-1.98	-1.16
		β_2	-0.55	0.0028	0.15	-0.87	-0.27
		τ	0.02	0.0002	0.01	0.01	0.03
	Helmeted Guineafowl	α_0	-0.10	0.0001	0.02	-0.15	-0.05
		α_1	0.80	0.0001	0.03	0.74	0.85
		β_0	1.85	0.0029	0.25	1.40	2.40
		β_1	-0.36	0.0005	0.09	-0.54	-0.20
		β_2	-0.36	0.0004	0.10	-0.56	-0.17
		τ	0.03	0.0002	0.01	0.01	0.05

The posterior summaries for some of the parameters of the nonspatial and spatial model are displayed in Table 11. The fixed regression effects of all of the parameters (for both data sets) were statistically different from zero since none of the 95% highest density credibility interval of the parameters contained zero. In all cases the regression effect for *nspp* were positive (as expected) while the regression effects of the occupancy effects were negative. The detection regression effects for both model types (for the respective species) were identical. The regression effects for the occupancy process for the Cape Weaver was

significantly different for the two model types while the same regression effects for the helmeted guineafowl were identical for both model types (except for the intercept). The posterior distribution of the spatial standard deviation parameter ($\sigma = 1/\sqrt{\tau}$) indicates that the spatial process **does significantly contribute** to the variability of the occupancy process across South Africa. The 95% posterior highest density credibility interval for σ are [5.56 , 10.59] and [4.26 , 8.42] for the Cape weaver and the helmeted guineafowl data sets respectively.

Figure 18 ((a) and (c)) displays the estimated occupancy probabilities ($Pr(z_i = 1|.)$) across South Africa estimated using *Rcppocc* for the Cape weaver and helmeted guineafowl data set respectively. The figures illustrate that there is a high probability that the Cape weaver occupies coastal regions throughout South Africa and low occupancy probability (close to zero) in the interior areas of South Africa. In contrast, the helmeted guineafowl has very high occupancy probabilities in most regions of South Africa except for the North West regions of South Africa. Figure 18 ((b) and (d)) displays the difference between the estimated occupancy probabilities obtained when using *Rcppocc* and *stocc* respectively ('*Rcppocc-stocc*'). The figures illustrate that we obtain similar estimates of the mean occupancy probabilities when using either estimation method with small discrepancies at the majority of the grid cells across South Africa.

4.4 Discussion and Conclusions

Through several studies, Bayesian methods have been developed to undertake occupancy models. They however either use probit link functions to model the detection and occupancy processes of the model; use general Bayesian analysis software such as *JAGS*, *WinBUGS*, *OpenBUGS* or *Stan* to undertake their analysis or make use of the Metropolis-Hastings algorithm to sample from the parameters of the model. We develop a Gibbs sampling algorithm to obtain posterior samples of the parameters of a restricted spatial regression (RSR) occupancy model and demonstrate that the method has a larger expected sampling rate (ESR) and faster run-times when compared to previous Bayesian methods used in the literature to date.

Similar to Broms et al. (2014) and Johnson et al. (2013) we show that the ICAR model produced posterior samples with significantly larger autocorrelations than the RSR model when using *stocc*. As an example, the autocorrelations of the occupancy regression effects as well as the spatial precision parameter (of both data sets) had autocorrelations in excess of 0.7 at lag 500 indicating that the posterior chain of the model mixed poorly for those parameters of the ICAR model. Additionally, the run-times of the ICAR model were approximately 5 times longer than the run-times of the RSR model and thus we do not recommend its use when fitting a spatial occupancy model.

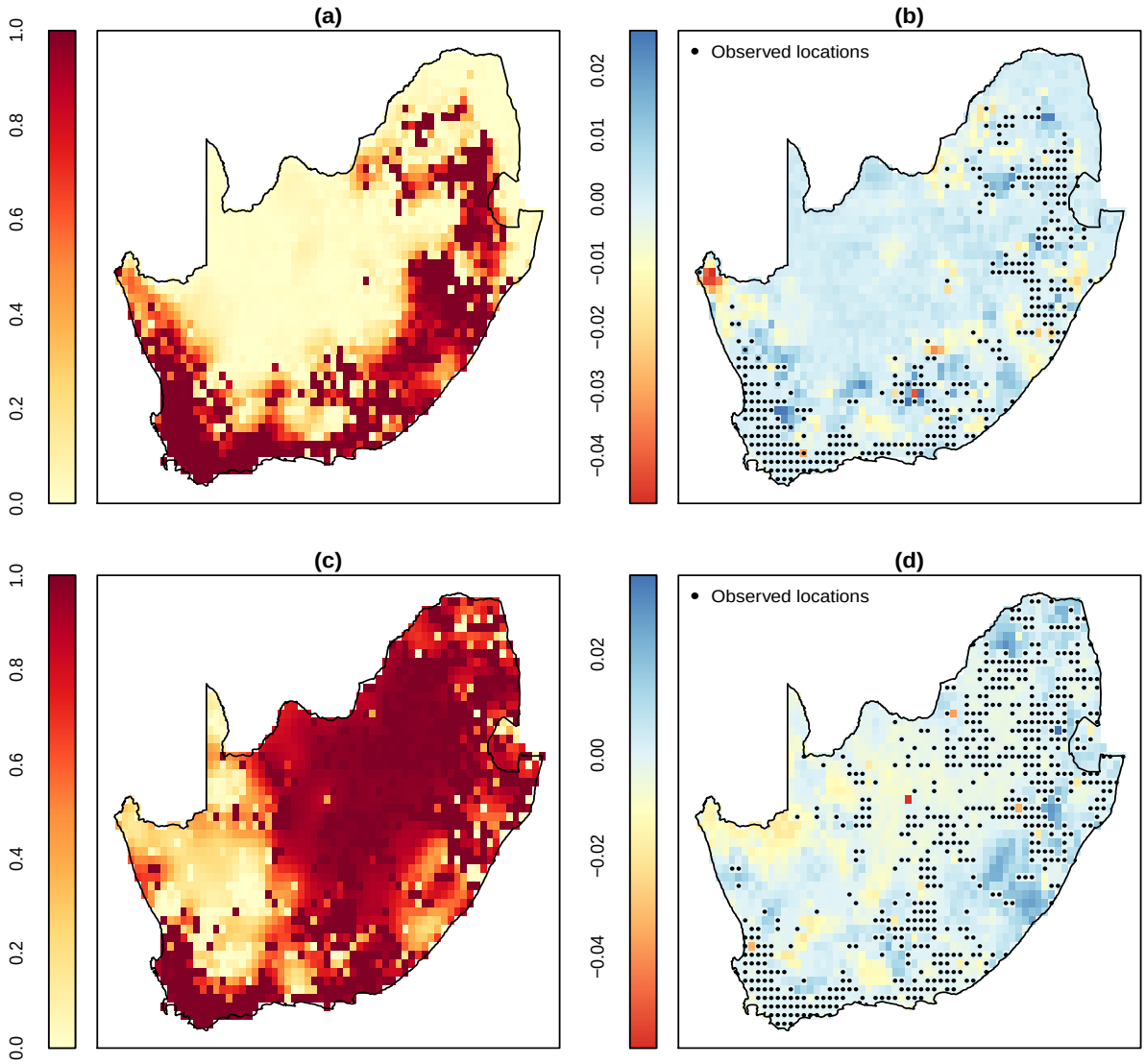

Figure 18. Estimated occupancy probability for the Cape weaver and helmeted guineafowl estimated using *Rcppocc* ((a) and (c)). The difference between the estimated occupancy probabilities obtained when using *Rcppocc* and *stocc* for the Cape weaver and helmeted guineafowl respectively ((b) and (d)). The grid-cells where the species have been detected at least once are displayed in (b) and (d).

Based on the two data sets, we observed that the new algorithm not only ran faster (approximately 35%) than the Gibbs sampler implemented in *stocc*, it also generated expected sample size (ESS) statistics between $2-6$ times larger than those obtained using *stocc*. The main reason for the time difference is that *stocc* has been coded using R while *Rcppocc* uses Rcpp and RcppArmadillo to undertake all matrix computations. *Stan* uses compiled C++ code to implement the *no-U-turn* Hamiltonian Monte Carlo algorithm and generated ESS statistics between $2-7$ times larger than those obtained using *Rcppocc*. In many applications *Stan* has been shown to be much faster than *JAGS* although at present *Stan* has run-times that are approximately $7-10$ times slower than *Rcppocc* when fitting spatial occupancy models. The opportunity thus exists to develop suitable *Stan* (or

NIMBLE) code that can fit spatial occupancy models in a shorter period of time.

Acknowledgements

This research was partially supported by two South African National Research Foundation grants namely 99385 (Clark) and 81685 (Altwegg). The financial assistance of the NRF towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

The authors have no conflict of interests to declare.

Allan Clark would also like to acknowledge the help of Andrew D. Crosby, a Post doctoral Fellow at the Boreal Avian Modelling Project, Department of Biological Sciences, University of Alberta. He shared code (with Allan Clark) on how to fit the single season occupancy model using *Stan* via email correspondence.

The authors would also like to thank Prof Linda Haines (University of Cape Town) for reading the initial manuscript and providing helpful comments.

Author Contribution

Below, Allan Ernest Clark is denoted as 'AEC', while Res Altwegg is denoted as 'RA'. AEC' and 'RA' conceived and designed the paper. 'AEC' analysed the data. 'AEC' wrote and, 'AEC' and 'RA' reviewed the paper. 'AEC' designed and coded the software used in the analysis. 'AEC' wrote computer code used to perform all analysis.

Appendix

Supporting information

The following sections contain Appendices to the manuscript.

4.A Gibbs algorithms to undertake logistic regression

A number of Gibbs sampling algorithms have been developed to obtain samples from the posterior distribution of the parameters of a **logistic regression model**. Holmes and Held (2006) represent the logistic distribution as a normal-mixture that utilizes a Kolmogorov-Smirnov distribution (Andrews and Mallows, 1974) as the mixing distribution while Frühwirth-Schnatter and Frühwirth (2007) and Frühwirth-Schnatter and Frühwirth (2010) both utilize the random-utility formulation of Mcfadden (1974) to develop their RUM (random utility method) and dRUM (difference of random utility method) algorithms. Both methods use *various approximations* as part of their algorithms. With the use of Pólya-Gamma random variables[3], Polson et al. (2013) developed a simple sampling scheme that generates samples from the correct posterior distribution without the aid of additional hierarchical levels or approximations. They evaluated the mixing properties of the above methods and found that **their method** performed best although Frühwirth-Schnatter and Frühwirth (2010) appeared to be a good alternative algorithm. Below we briefly discuss both of these formulations and explain how they could be used to obtain posterior samples for the parameters of the single season occupancy (SSO) model.

Following Frühwirth-Schnatter and Frühwirth (2010), the logistic regression model can be formulated using unobserved latent random variables z_i^{u*}, for $i = 1, \ldots, n_s$ where n_s is the number of surveyed sites. These random variables are modelled using explanatory variables with design matrix $\boldsymbol{X}$, while the error of the relationship is modelled using a Logistic distribution[4]. We thus have

$$z_i^{u*} = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, \ \epsilon_i \sim \text{Logistic}(0,1) \text{ with } y_i = I_{\{z_i^{u*}>0\}}, \text{ for all } i = 1, \ldots, n_s.$$

The novelty of the method lies in the approximation of $p(\epsilon)$ using a linear Gaussian mixture distribution with **known weights**, w_r^* and variances, s_r^2 for $r = 1, 2, 3$. Since the Logistic distribution is symmetric about 0, the means of the Gaussian mixture distribution are assumed to be 0 as well. The logistic model is now reformulated using the following hierarchical framework,

$$z_i^{u*} = \boldsymbol{x}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{4.4}$$

[3]"A random variable X has a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \Re$, denoted $X \sim \text{PG}(b, c)$, if

$$X \overset{D}{=} \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{g_k}{(k-0.5)^2 + c^2/(4\pi^2)},$$

where $g_k \sim \mathcal{G}(b, 1)$ are independent gamma random variables, and where $\overset{D}{=}$ indicates equality in distribution" (Polson et al., 2013).

[4]The probability density function of a Logistic distribution with parameters $a \in \Re$ and $b > 0$ is $f(x) = \dfrac{e^{-(x-a)/b}}{b(1+e^{-(x-a)/b})^2}$ (Balakrishnan, 1991).

$$\epsilon_i | r_i \sim \mathcal{N}(0, \omega_i) \text{ with } \omega_i = s^2_{r_i} \text{ and} \tag{4.5}$$

$$r_i \sim \text{Multinomial}(w_1, w_2, w_3), \text{ for all } i = 1, \ldots, n_s. \tag{4.6}$$

The use of the linear Gaussian approximation allows one to perform weighted linear regression to obtain the posterior distribution of the regression coefficients conditional on knowing the $\boldsymbol{z}^{u*}$ vector as well as $\boldsymbol{w} = [w_1, w_2, w_3]^T$ and $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$. The conditional posterior distributions of z_i^{u*} and r_i are easily obtained using equations (4.4-4.6) in the framework above.

Polson et al. (2013) show that posterior samples of $\boldsymbol{\beta}$ can be obtained by sampling from the following conditional distributions, in turn,

$$\begin{aligned} \omega_i | \boldsymbol{\beta} &\sim \text{PG}(1, \boldsymbol{x}_i^T \boldsymbol{\beta}), \text{ for all } i = 1, \ldots, n_s \\ \boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{\omega} &\sim \mathcal{N}(\boldsymbol{\mu}, \ \boldsymbol{v}) \end{aligned}$$

where the 'PG' notation represents a Pólya-Gamma random variable, $\boldsymbol{\mu}$ and $\boldsymbol{v}$ are the mean vector and covariance matrix of a multivariate Gaussian distribution.

Below we briefly investigate how both of these methods could be used to sample from the posterior distribution of the parameters of a SSO model, with the aim of deciding which of the two formulations (see Appendix 4.B and Appendix 4.C) should be used to develop a Gibbs sampling algorithm for spatial occupancy models.

4.A.1 A Gibbs algorithm for the SSO model

Since an occupancy model has regression effects in both the occupancy and detection process, two sets of additional random variables are introduced into the sampling algorithm. The dRUM method (Algorithm 1 in Table 12) thus require posterior samples for $\boldsymbol{z}^{u*}$, $\boldsymbol{r}$, $\boldsymbol{y}^{u*}$ and $\boldsymbol{r}^*$, where the first two relate to the occupancy regression effects and the last two relates to the detection regression effects. Algorithm 2 (see Table 12) can be used to generate samples from two sets of Pólya-Gamma random variables, $\boldsymbol{\omega_\alpha}$ and $\boldsymbol{\omega_\beta}$, in order to obtain posterior samples for the regression effects of an SSO model.

Table 12 displays a summary of the sampling scheme required to sample from the posterior distribution of the parameters associated with the SSO model for the two Gibbs sampling algorithms considered in the paper. It should be immediately apparent that Algorithm 1 requires **twice** as many additional parameters to sample from than Algorithm 2. Algorithm 2 however requires samples from a Pólya-Gamma distribution which requires the use of an acceptance-rejection algorithm (Robert and Casella, 1999). Both algorithms have been coded using the R programming language (R Core Team, 2014) in conjunction with *RcppArmadillo* (Sanderson and Curtin, 2016) and *Rcpp* (Eddelbuettel and Francois, 2011).

The Pólya-Gamma random variables were generated by amending the Rcpp code in the R package *Binomlogit* (Windle et al., 2013). The exact details related to the algorithms can be found in Appendix 4.B and Appendix 4.C.

Table 12. Two Gibbs algorithms for a SSO model. (See Appendix 4.B and 4.C for the details pertaining to the parameter matrices of the conditional posterior distributions.)

Algorithm 1. (DRUM)

Set starting values for $\boldsymbol{\beta}$ and $\boldsymbol{z}$.
for (iterations= $1, \ldots,$ simulation runs) **do**
  for $(i = 1, \ldots, n_s)$ **do**
    - Generate $z_i^{u*} \sim [z_i^{u*} | \boldsymbol{z}, \boldsymbol{\beta}]$.
    - Generate $r_i \sim [r_i | z_i^{u*}, \boldsymbol{\beta}]$.
  end for
  - Generate $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta,\ \boldsymbol{\Sigma}_\beta)$.
  for $((ij) = \{z_i = 1\})$ **do**
    - Generate $y_{ij}^{u*} \sim [y_{ij}^{u*} | \boldsymbol{y}, \boldsymbol{\alpha}]$.
    - Generate $r_{ij}^{*} \sim [r_{ij}^{*} | \boldsymbol{y}, \boldsymbol{y}^{u,*}]$.
  end for
  - Generate $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$.
  - Generate $z_i \sim [z_i | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{y}]$.
end for

Algorithm 2. (PG)

Set starting values for $\boldsymbol{\beta}$ and $\boldsymbol{z}$.
for (iterations= $1, \ldots,$ simulation runs) **do**
  for $(i = 1, \ldots, n_s)$ **do**
    - Generate $\omega_{i,\beta} \sim \text{PG}(1, \boldsymbol{x}_i^T \boldsymbol{\beta})$.
  end for
  - Generate $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_\beta,\ \boldsymbol{\Sigma}_\beta)$.
  for $((ij) = \{z_i = 1\})$ **do**
    - Generate $\omega_{ij,\alpha} \sim \text{PG}(1, \boldsymbol{w}_{ij}^T \boldsymbol{\alpha})$.
  end for
  - Generate $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$.
  - Generate $z_i \sim [z_i | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{y}]$.
end for

4.A.2 Simulation Study

As noted above, Polson et al. (2013) found that the use of Pólya-Gamma random variables lead to MCMC chains that **mixed faster** than alternate algorithms when undertaking **logistic regression**. Below we investigate whether or not this observation holds when applying Algorithm 2 to simulated detection-nondetection data. We utilize the effective sample size (ESS) and the effective sampling rate (ESR) (effective sample size per second of run-time) to compare the efficiency of MCMC chains produced using *JAGS*, Algorithm 1, Algorithm 2 and *Stan*. The effective sample size for the i^{th} parameter in the model is defined as

$$ESS_i = \frac{M}{1 + 2\sum_{j=1}^{k} \rho_i(j)},$$

where M is the number of post-burn-in samples, and $\rho_i(j)$ is the j^{th} lagged autocorrelation of parameter i (Holmes and Held, 2006). We use the *coda* package (Plummer et al., 2006) to estimate ESS_i and use boxplots of the median (*across all regression effects*) effective

sample size and median effective sampling rate to draw conclusions regarding the efficiency of the four sampling methods.

The simulation settings discussed below are very similar to that of Clark et al. (2016). We consider the following simulation settings. The number of sites (n_s) are set to 50 and 100 while the number of surveys to each site (J) are set to 3 and 5 respectively. The following combinations of the regression coefficients are used: 1. $\boldsymbol{\alpha} = [0,\ 1.75]^T$, $\boldsymbol{\beta} = [-1.85,\ 2.5]^T$; 2. $\boldsymbol{\alpha} = [0,\ 1.75]^T$, $\boldsymbol{\beta} = [-0.1,\ 2.5]^T$; 3. $\boldsymbol{\alpha} = [1.35,\ 1.75]^T$ $\boldsymbol{\beta} = [-1.85,\ 2.5]^T$ and 4. $\boldsymbol{\alpha} = [1.35,\ 1.75]^T$, $\boldsymbol{\beta} = [-0.1,\ 2.5]^T$. These parameter values ensure an approximate average detection and occupancy probability among the sites of (0.5, 0.3), (0.5, 0.5), (0.7, 0.3) and (0.7, 0.5) respectively. In order to assess how well these algorithms perform for larger sample sizes we also consider the following cases, $n_s = 500$ with $J = 5$ and $J = 10$ using the first two parameter settings described above.

"The occupancy regression covariate was obtained by standardizing a Uniform $(-2, 2)$ random variable while the detection covariate was obtained by standardizing a Uniform$(-5, 5)$ random variable. Each of these variables were transformed to have a zero mean and a standard deviation of one. The following parameter vectors were used to specify the prior distribution of the parameters: $\boldsymbol{\mu}_{\boldsymbol{i}}^0 = [0, 0]^T$, $\boldsymbol{\Sigma}_{\boldsymbol{i}}^0 = \text{diag}[1000, 1000]$ for $\boldsymbol{i} = \boldsymbol{\alpha}, \boldsymbol{\beta}$." (Clark et al., 2016)

Each simulation setting was replicated 500 times. All calculations were undertaken using R 3.4.1 (R Core Team (2014)). MCMC sampling was undertaken using the R packages *jagsUI* (Kellner, 2014) and *rstan* in combination with *JAGS* 4.2.0 (Plummer, 2003) and *Stan* 2.17.3 (Carpenter et al., 2017) while Algorithm 1 and Algorithm 2 were performed using the authors' code. A copy of the R package created to undertake the posterior sampling from the SSO model can be found at `https://github.com/AllanClark/Rcppocc`. Note that the **default settings** for the *Stan* function, *sampling*, were used in order to undertake sampling when using *Stan*. 20 000 posterior samples was obtained for each MCMC simulation. The first 10 000 samples were discarded as burn-in samples while the remaining samples were retained. Prior experimentation using the MCMC algorithm indicated that the Markov chains would converge to the stationary distributions using this number of posterior samples. The posterior samples were not thinned (Link and Eaton, 2012).

4.A.3 Results: Simulation Study

Based on the simulation study, we conclude that Algorithm 2 had the fastest **run-times** across all scenarios considered. *JAGS* and Stan had similar run-times for $n_s \leq 100$ with median run-times being between $6 - 12$ times slower than Algorithm 2. The *JAGS* algorithm does not scale well and took $40 - 70$ times longer to complete than Algorithm 2 when $n_s = 500$ and comparisons are made using the median run-times across of all

regression effects. We found that Algorithm 1 performed well relative to Algorithm 2 except for $n_s = 500$ and in general we can conclude that for small sample sizes the run-times for *JAGS* and *Stan* are similar although *Stan* outperforms *JAGS* for large sample sizes (see Figure 19). A similar result was found by Monnahan et al. (2017) when investigating various population ecology models.

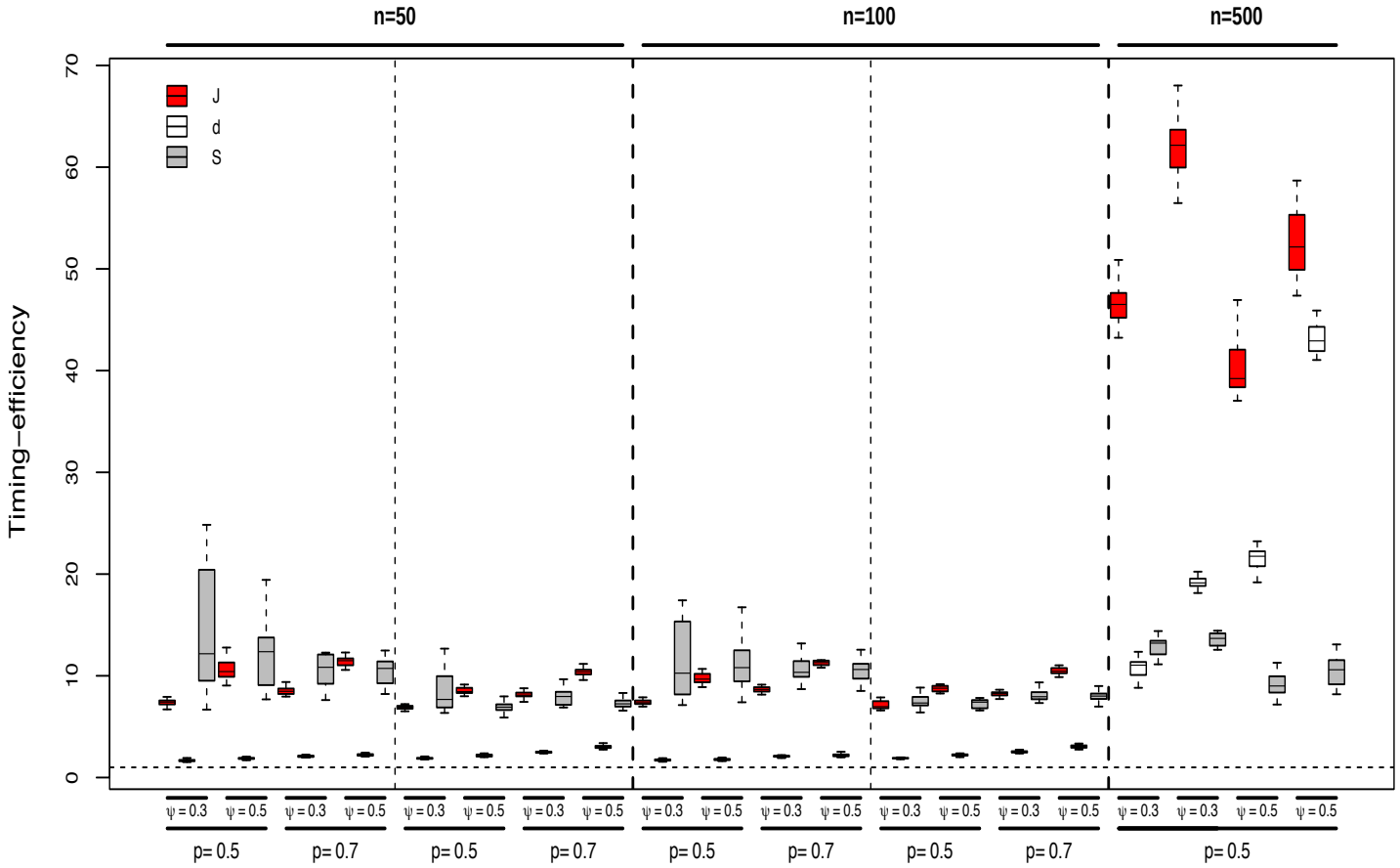

Figure 19. Trimmed Box plots of the relative efficiency (run-time of an algorithm in seconds divided by the run-time of Algorithm 2 in seconds) of *JAGS* (J), Algorithm 1 (d) and *Stan* (S) when generating 20 000 posterior samples from the posterior distribution of the SSO model for all simulation scenarios considered. The average occupancy probabilities and conditional detection probabilities are denoted as ψ and p. The horizontal dashed line represents a time-efficiency of 1.

We found that the median ESS obtained when using *Stan* is significantly larger than those obtained when using the other three algorithms. *JAGS* and Algorithm 1 produced similar ESS's while Algorithm 2 produced larger ESS's than *JAGS* and Algorithm 1 in general (see Figures 20 and 21).

The median ESR obtained when using Algorithm 2 is **generally much larger** than the median ESR obtained when using the other three algorithms and occurs largely due to the faster run-times of Algorithm 2 (see Figures 20 and 21). For $n_s \leq 100$, the ESR for Algorithm 2 is $3-12$ times larger than those obtained when using Stan; $3-5$ times larger than those obtained when using Algorithm 1 and $10-120$ times larger than those obtained when using JAGS (*when one does not use the median regression effects to combine the results).* When $n_s = 500$, Algorithm 2 performs significantly better than all other algorithms with the ESR being $3-10$ times larger than those obtained when using Stan, $15-60$ times larger than those obtained when using Algorithm 1 and $30-450$ times larger than those obtained when using *JAGS*.

The above observations are broadly consistent across all scenarios considered. Regarding large data sizes; we can conclude that this negatively effects ESR although Algorithm 2 clearly outperforms the other three sampling algorithms. The poor performance of *JAGS* was expected and occurs since the software has been programmed to implement Bayesian analysis in many different situations. The software has not been written or tailored for speed and efficiency whereas Algorithms 1 and 2 were custom built to undertake the Bayesian analysis of SSO models using Rcpp. Recall also that Algorithm 1 requires the sampling of twice as many additional parameters than Algorithm 2. Algorithm 2 performed surprisingly well relative to *Stan*. *Stan* does not allow for sampling of discrete random variables and in an occupancy modelling context, we **marginalised** all true occupancy variables from the posterior distribution which results in much less parameters being sampled.

Based on the **above** simulation results we propose that Pólya-Gamma random variables (instead of the dRUM formulation) be used to develop a Gibbs sampling algorithm for a spatial occupancy model (see Appendix 4.D below for such an algorithm). We have not extensively compared the two algorithms when fitting spatial occupancy models although believe that similar conclusions as was found above would be obtained (see Figure 22).

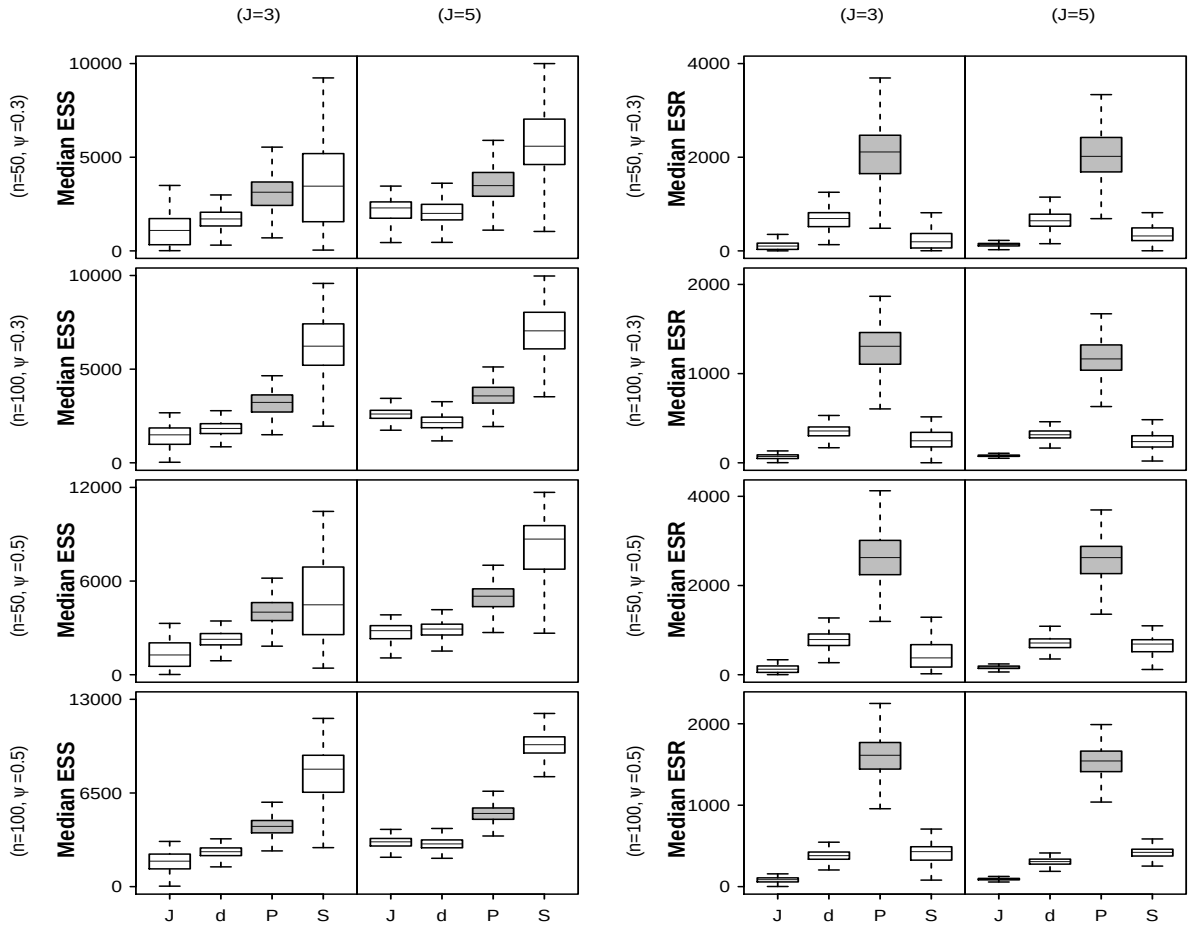

Figure 20. Trimmed Box plots of the median effective sample size (ESS) and median effective sampling rate (ESR) associated with MCMC chains produced using *JAGS* (J), Algorithm 1 (d), Algorithm 2 (P) and *Stan* (S) when fitting a nonspatial occupancy model. The number of sites are set to $n_s = 50$ and $n_s = 100$ while the number of surveys at each site are set to $K = 3$ and $K = 5$. The detection probability for these scenarios are approximately 0.5.

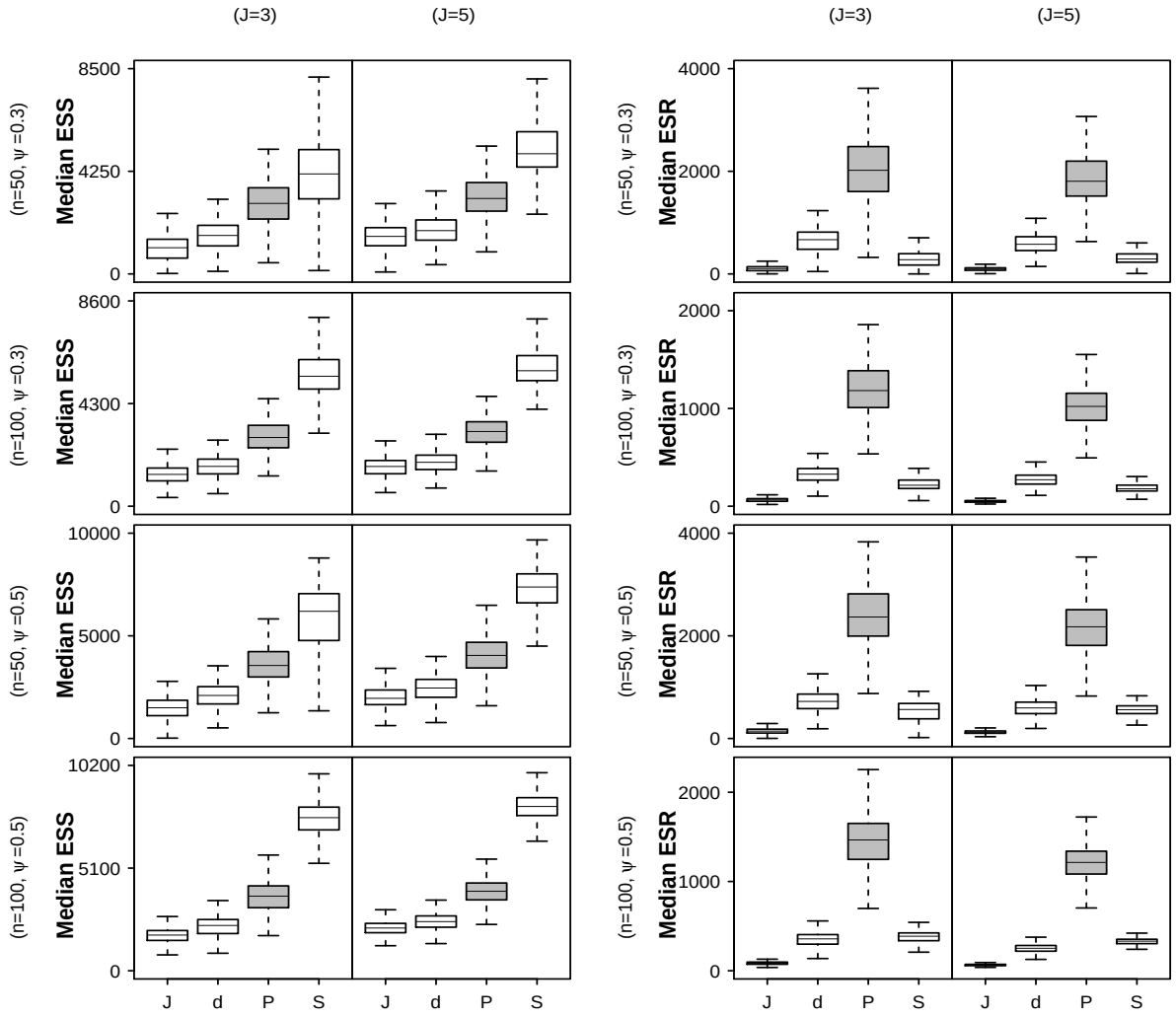

Figure 21. Trimmed Box plots of the median effective sample size (ESS) and median effective sampling rate (ESR) associated with MCMC chains produced using *JAGS* (J), Algorithm 1 (d), Algorithm 2 (P) and *Stan* (S). The number of sites are set to $n_s = 50$ and $n_s = 100$ while the number of surveys to each site are set to $K = 3$ and $K = 5$. The detection probability for these scenarios are approximately 0.7.

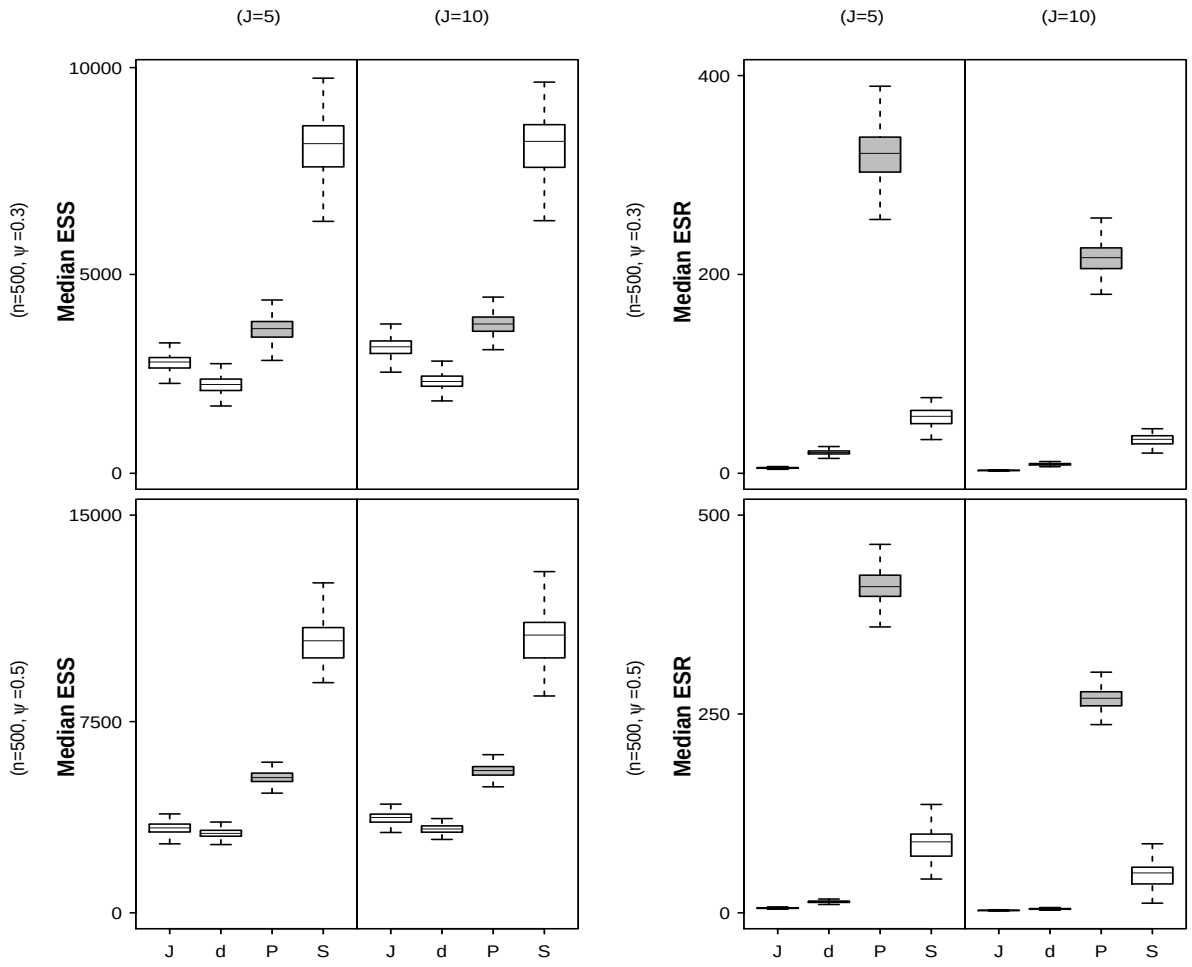

Figure 22. Box plots of the median effective sample size (ESS) and median effective sampling rate (ESR) associated with MCMC chains produced using *JAGS* (J), Algorithm 1 (d), Algorithm 2 (P) and *Stan* (S). The number of sites are set to $n_s = 500$ while the number of surveys to each site are set to $K = 5$ and $K = 10$. The detection probability for these scenarios are approximately 0.5.

4.B Gibbs algorithm for the SSO model using dRUM.

Below we provide the full conditional distributions necessary to implement a Gibbs sampler for the SSO model when using the dRUM formulation. Here we assume that the prior distributions are $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)$ and $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)$. In addition, let $\boldsymbol{X}$ be a design matrix for the occupancy process with rows $\boldsymbol{x}_i^T$ and $\boldsymbol{W}$ be the design matrix for the detection process with rows $\boldsymbol{w}_{ij}^T$. If these prior distributions are adopted the Gibbs sampler proceeds as follows.

1. Set starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{z}$.

2. Cycle:

   2.1 Generate the latent utilities $z_i^{u*},\ i = 1, \ldots, n_s$ conditional on $\boldsymbol{\beta}$ as

   $$z_i^{u*} = \ln(\lambda_1 U_i + z_i) - \ln(1 - U_i + \lambda_i(1 - z_i)),$$

   where U_i are independent Uniform random variables and $\ln \lambda_i = \boldsymbol{x}_i^T \boldsymbol{\beta}$.

   2.2 Generate the indicator variable r_i conditional on z_i^{u*} from

   $$\Pr(r_i = j | z_i^{u*}, \boldsymbol{\beta}) \propto \frac{w_j}{s_j} \exp\left(-\frac{1}{2}\left(\frac{z_i^{u*} - \ln \lambda_i}{s_j}\right)^2\right),$$

   and set $\omega_i = s_{r_i}^2$.

   2.3 Sample from $[\boldsymbol{\beta}|.]$ where $\boldsymbol{R} = (r_1, \ldots, r_{n_s})$, $\boldsymbol{\beta}|. \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$ where

   $$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\beta}} &= \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{X}^T \boldsymbol{S}_{\boldsymbol{\beta}}^{-1} \boldsymbol{z}^{u*} \text{ and} \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}} &= \left((\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)^{-1} + \boldsymbol{X}^T \boldsymbol{S}_{\boldsymbol{\beta}}^{-1} \boldsymbol{X}\right)^{-1}. \end{aligned}$$

   $\boldsymbol{S}_{\boldsymbol{\beta}}$ is a diagonal matrix with diagonal entries $\omega_{1,\boldsymbol{\beta}}, \ldots, \omega_{n_s,\boldsymbol{\beta}}$.

   2.4 Generate the latent utilities $y_{ij}^{u*},\ i = 1, \ldots, n^*$ conditional on $\boldsymbol{\alpha}$ as

   $$y_{ij}^{u*} = \ln(\lambda_{ij} U_i^* + y_{ij}) - \ln(1 - U_i^* + \lambda_{ij}(1 - y_{ij})),$$

   where U_i^* are independent Uniform random variables and $\ln \lambda_{ij} = \boldsymbol{w}_{ij}^T \boldsymbol{\alpha}$. Here the indices (ij) index relates to those sites with $z_i = 1$.

   2.5 Generate the indicator variable r_{ij}^* conditional on y_{ij}^{u*} from

   $$\Pr(r_{ij}^* = j | y_{ij}^{u*}, \boldsymbol{\alpha}) \propto \frac{w_j}{s_j} \exp\left(-\frac{1}{2}\left(\frac{y_{ij}^{u*} - \ln \lambda_{ij}}{s_j}\right)^2\right),$$

   and set $\omega_{ij} = s_{r_i}^2$.

2.6 Sample from $[\boldsymbol{\alpha}|.]$ where $\boldsymbol{R}^* = (r^*_{1,1}, \ldots, r^*_{n^*,K_{n^*}})$, $\boldsymbol{\alpha}|. \sim \mathcal{N}(\boldsymbol{\mu_\alpha},\ \boldsymbol{\Sigma_\alpha})$ where

$$\begin{aligned}\boldsymbol{\mu_\alpha} &= \boldsymbol{\Sigma_\alpha}\tilde{\boldsymbol{W}}^T\boldsymbol{S}_{\boldsymbol{\alpha}}^{-1}\boldsymbol{y}^{u*} \text{ and} \\ \boldsymbol{\Sigma_\alpha} &= \left((\boldsymbol{\Sigma}^0_{\boldsymbol{\alpha}})^{-1} + \tilde{\boldsymbol{W}}^T\boldsymbol{S}_{\boldsymbol{\alpha}}^{-1}\tilde{\boldsymbol{W}}\right)^{-1}.\end{aligned}$$

$\boldsymbol{S_\alpha}$ is a diagonal matrix with diagonal entries $\omega_{1,1,\boldsymbol{\alpha}}, \ldots, \omega_{n^*,K_{n^*},\boldsymbol{\alpha}}$. The matrix $\tilde{\boldsymbol{W}}$ relates to those entries in $\boldsymbol{W}$ associated with $z_i = 1$.

2.7 If the species is observed at location i, then $z_i = 1|. \sim$ Bernoulli(1) but if the species is not observed at location i, then

$$z_i = 1|. \sim \text{Bernoulli}\left(\frac{\psi_i \prod_j (1 - p_{ij})}{1 - \psi_i + \psi_i \prod_j (1 - p_{ij})}\right)$$

where ψ_i and p_{ij} are the occupancy and conditional detection probability of a species associated with each site and survey occasion.

4.C Gibbs algorithm for the SSO model using Pólya-Gamma random variables.

Below we provide the full conditional distributions necessary to implement a Gibbs sampler for the SSO model when using Pólya-Gamma random variables. Here we assume that the prior distributions are $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0,\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)$ and $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}^0,\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)$. Denote the prior precision matrices of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^0$ and $\boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0$.

1. Set starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{z}$.

2. Cycle:

   2.1 Generate $\omega_{i,\boldsymbol{\beta}}|. \sim \text{PG}(1, \boldsymbol{x}_i^T\boldsymbol{\beta})$.

   2.2 Sample from $[\boldsymbol{\beta}|.]$ where $\boldsymbol{\beta}|. \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$ with,

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left(\boldsymbol{X}^T(\boldsymbol{z} - 0.5\mathbf{1}_n) + \boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0\boldsymbol{\mu}_{\boldsymbol{\beta}}^0\right) \text{ and}$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \left(\boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0 + \boldsymbol{X}^T\boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{X}\right)^{-1}.$$

   $\boldsymbol{S}_{\boldsymbol{\beta}}$ is a diagonal matrix with diagonal entries $\omega_{1,\boldsymbol{\beta}}, \ldots, \omega_{n_s,\boldsymbol{\beta}}$.

   2.3 Generate $\omega_{ij,\boldsymbol{\alpha}}|. \sim \text{PG}(1, \boldsymbol{w}_{ij}^T\boldsymbol{\alpha})$.

   2.4 Sample from $[\boldsymbol{\alpha}|.]$ where $\boldsymbol{\alpha}|. \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ with

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}\left(\tilde{\boldsymbol{W}}^T(\tilde{\boldsymbol{y}} - 0.5\mathbf{1}_{n^*}) + \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^0\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0\right) \text{ and}$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \left(\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^0 + \tilde{\boldsymbol{W}}^T\boldsymbol{S}_{\boldsymbol{\alpha}}\tilde{\boldsymbol{W}}\right)^{-1}.$$

   $\boldsymbol{S}_{\boldsymbol{\alpha}}$ is a diagonal matrix with diagonal entries $\omega_{1,1,\boldsymbol{\alpha}}, \ldots, \omega_{n^*,K_{n^*},\boldsymbol{\alpha}}$. The matrices $\tilde{\boldsymbol{W}}$ and $\tilde{\boldsymbol{y}}$ relates to those entries in $\boldsymbol{W}$ and $\tilde{\boldsymbol{y}}$ associated with $z_i = 1$. The $\tilde{\boldsymbol{y}}$ vector is constructed by column stacking the entries of the observed detection-nondetection data $\boldsymbol{y}$.

   2.5 Sampling from $[z_i|.]$ is the same as in Appendix 4.B.

4.D Gibbs algorithm for a spatial occupancy model using Pólya-Gamma random variables.

Denote the true occupancy variable as $\boldsymbol{z} = [\boldsymbol{z}^{(s)}, \boldsymbol{z}^{(s)'}]^T$ where $\boldsymbol{z}^{(s)}$ and $\boldsymbol{z}^{(s)'}$ represents the true occupancy variables at sites that were **surveyed** and **not surveyed** respectively. n represents the number of sites in the study (i.e. $n = \text{length}(\boldsymbol{z})$) while n_s is the number of sites surveyed. Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}^{(s)} \\ \boldsymbol{X}^{(s)'} \end{bmatrix} \text{ and } \boldsymbol{K} = \begin{bmatrix} \boldsymbol{K}^{(s)} \\ \boldsymbol{K}^{(s)'} \end{bmatrix} \tag{4.7}$$

where both $\boldsymbol{X}^{(s)}$ and $\boldsymbol{K}^{(s)}$ have n_s rows.

Below we provide the full conditional distributions necessary for implementing a Gibbs sampler for a restricted spatial regression (RSR) occupancy model using Pólya-Gamma random variables. Here we assume that the prior distributions are $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^0,\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^0)$, $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}^0,\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^0)$, $\boldsymbol{\theta}|\tau \sim \mathcal{N}\left(\boldsymbol{0},\ \frac{1}{\tau}\boldsymbol{M}\right)$ and $\tau \sim \mathcal{G}(i_1,\ i_2)$. Denote the prior precision matrices of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as $\boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^0$ and $\boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0$.

1. Set starting values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and $\boldsymbol{z}^{(s)}$.

2. Cycle:

   2.1 Generate $\omega_{i,\boldsymbol{\beta}}|. \sim \text{PG}(1, \boldsymbol{x}_i^T\boldsymbol{\beta} + \boldsymbol{k}_i^T\boldsymbol{\theta}),$ for all $i = 1, \ldots, n_s$.

   2.2 Sample from $[\boldsymbol{\beta}|.]$ where $\boldsymbol{\beta}|. \sim \mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$ with

   $$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\beta}} &= \boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left(\boldsymbol{X}^{(s)T}(\boldsymbol{z}^{(s)} - 0.5\boldsymbol{1}_{n_s} - \boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{K}^{(s)}\boldsymbol{\theta}) + \boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0\boldsymbol{\mu}_{\boldsymbol{\beta}}^0\right) \text{ and} \\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}} &= \left(\boldsymbol{\Lambda}_{\boldsymbol{\beta}}^0 + \boldsymbol{X}^{(s)T}\boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{X}^{(s)}\right)^{-1}. \end{aligned}$$

   $\boldsymbol{S}_{\boldsymbol{\beta}}$ is a diagonal matrix with diagonal entries $\omega_{1,\boldsymbol{\beta}}, \ldots, \omega_{n_s,\boldsymbol{\beta}}$.

   2.3 Sample from $[\boldsymbol{\theta}|.]$ where $\boldsymbol{\theta}|. \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$ with

   $$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\theta}} &= \boldsymbol{\Sigma}_{\theta}\boldsymbol{K}^{(s)T}\left(\boldsymbol{z}^{(s)} - 0.5\boldsymbol{1}_{n_s} - \boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{X}^{(s)}\boldsymbol{\beta}\right) \\ \boldsymbol{\Sigma}_{\boldsymbol{\theta}} &= \left(\tau\boldsymbol{M}^{-1} + \boldsymbol{K}^{(s)T}\boldsymbol{S}_{\boldsymbol{\beta}}\boldsymbol{K}^{(s)}\right)^{-1}. \end{aligned}$$

   2.4 Generate $\tau|. \sim \mathcal{G}\left(\frac{r}{2} + i_1,\ \frac{\boldsymbol{\theta}^T\boldsymbol{M}^{-1}\boldsymbol{\theta}}{2} + i_2\right)$.

   2.5 Generate $\omega_{ij,\boldsymbol{\alpha}}|. \sim \text{PG}(1, \boldsymbol{w}_{ij}^T\boldsymbol{\alpha}),$ for all (ij) satisfying $\{z_i = 1\}$.

2.6 Sample from $[\boldsymbol{\alpha}|.]$ as done in Appendix 4.C.

2.7 (a) **For a surveyed site:** If the species is observed at location i, then $z_i|. \sim$ Bernoulli(1) but if the species is not observed at location i, then

$$z_i|. \sim \text{Bernoulli}\left(\frac{\psi_i \prod_j (1-p_{ij})}{1-\psi_i+\psi_i \prod_j (1-p_{ij})}\right)$$

where ψ_i and p_{ij} are the occupancy and conditional detection probability of a species associated with each site and survey occasion.

2.7 (b) **For an unsurveyed site:** $z_i|. \sim \text{Bernoulli}\left(\psi_i\right).$

4.D.1 A brief derivation of the conditional distribution of the occupancy regression effects.

Take note that the conditional posterior distributions of $\boldsymbol{\beta}$ rely on the observation that $\psi^{z_i}(1-\psi_i)^{1-z_i}$ can be rewritten as

$$\frac{\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta}\right)^{z_i}}{1+\exp\left(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta}\right)} = \frac{1}{2}\exp(\kappa_i(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta}))\int \exp\left(-\frac{\omega_{i,\beta}}{2}(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta})^2\right) p(\omega_{i,\beta}|1,0)d\omega_{i,\beta}$$

[5]where $p(\omega_{i,\beta}|1,0)$ is the probability density function of a Pólya-Gamma distribution with parameters 1 and 0. $\kappa_i = z_i - 0.5$ (Polson et al., 2013).

If we condition on $\boldsymbol{\omega}_\beta$, the conditional posterior distribution of the occupancy regression effects are proportional to

$$[\boldsymbol{\beta}|.] \propto \pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \exp(\kappa_i(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta}))\exp\left(-\frac{\omega_{i,\beta}}{2}(\boldsymbol{x}_i^T\boldsymbol{\beta}+\boldsymbol{k}_i^T\boldsymbol{\theta})^2\right)$$

and after completing the square in terms of $\boldsymbol{\beta}$ we obtain the conditional posterior distribution as reported in the above algorithm.

The conditional posterior distribution of $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ both use the same manipulation of the Bernoulli likelihood.

[5]The '$\frac{1}{2}$' was omitted in the published paper.

4.E Summary statistics of the climate covariates.

- The variables *GDD0* and *GDD5* relate to annual thermal sums above 0 and 5 degrees centigrade respectively.
- The variables *MTCO* and *MTWA* relate to the mean temperature of the coldest and warmest month respectively.
- The variable *AETPET* is the ratio of potential to realised evapotranspiration.
- The variables *DryInt* and *WetInt* relates to the intensity of the dry and wet season respectively.

Table 13. Lower triangle of the correlation matrix of the climate variables.

	GDD0	GDD5	MTCO	MTWA	AETPET	WetInt	DryInt
GDD0	1						
GDD5	1.00	1					
MTCO	**0.86**	**0.86**	1				
MTWA	**0.85**	**0.85**	0.50	1			
AETPET	-0.27	-0.27	0.05	**-0.61**	1		
WetInt	-0.24	-0.23	0.11	-0.51	**0.74**	1	
DryInt	-0.34	-0.34	-0.00	**-0.65**	**0.93**	**0.62**	1

Table 14. Variance inflation factors of the climate variables.

Variables	GDD0	GDD5	MTCO	MTWA	AETPET	WetInt	DryInt
VIF	3 393.35	3 371.0	25.85	32.65	16.74	4.44	10.40

Table 15. Summary statistics of the principal components of the climate variables.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	2.03	1.48	0.64	0.45	0.24	0.099	0.012
Cumulative Proportion	0.59	0.90	0.96	0.99	0.999	0.9999	1

4.F Plots of the climate variables and the associated principal components.

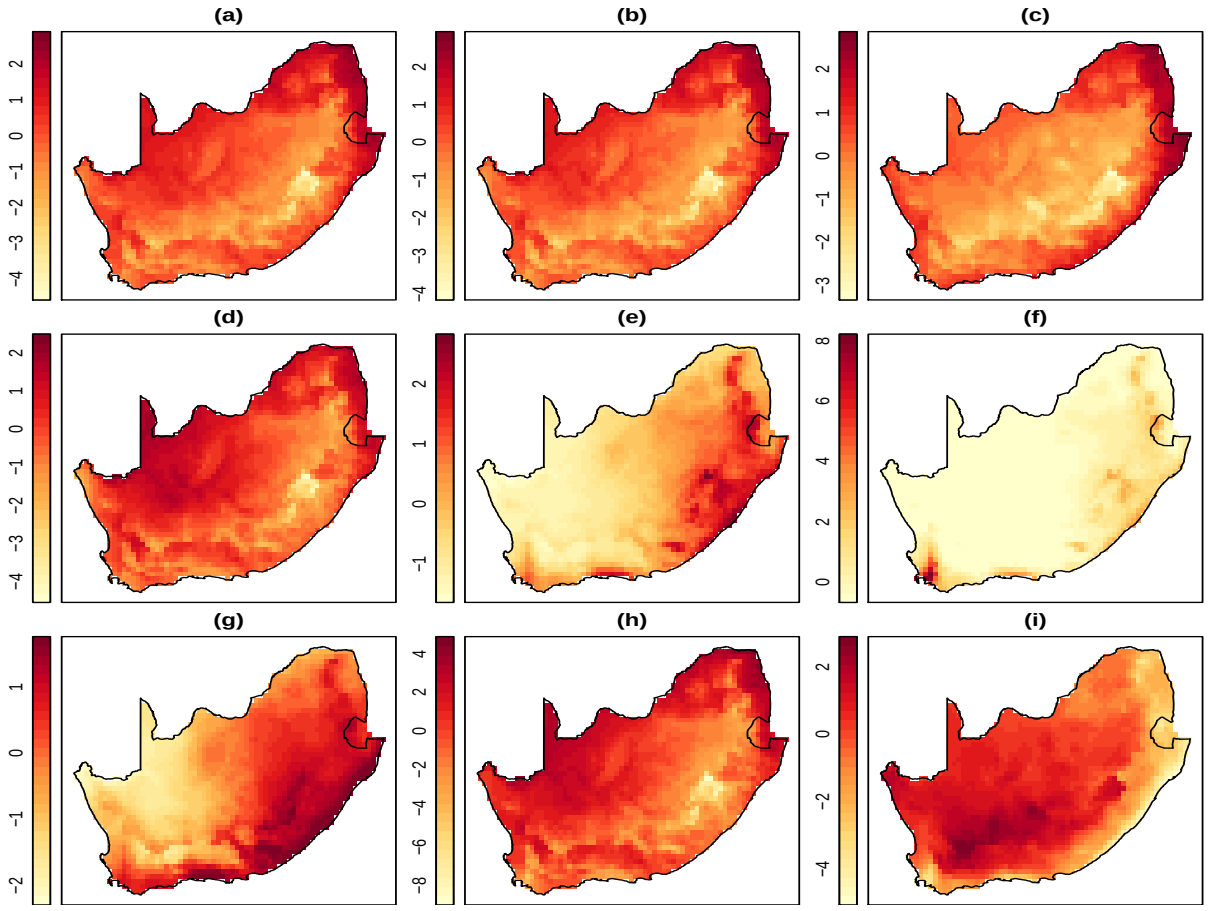

Figure 23. The climate variables as well as the first two principal components of the climate variables ((a) = GDD0, (b) = GDD5, (c)= MTCO, (d)= MTWA, (e) = AETPET, (f)= WetInt, (g) = DryInt, (h) = 1st principal component, (i) = 2nd principal component).

4.G Certain posterior distributions.

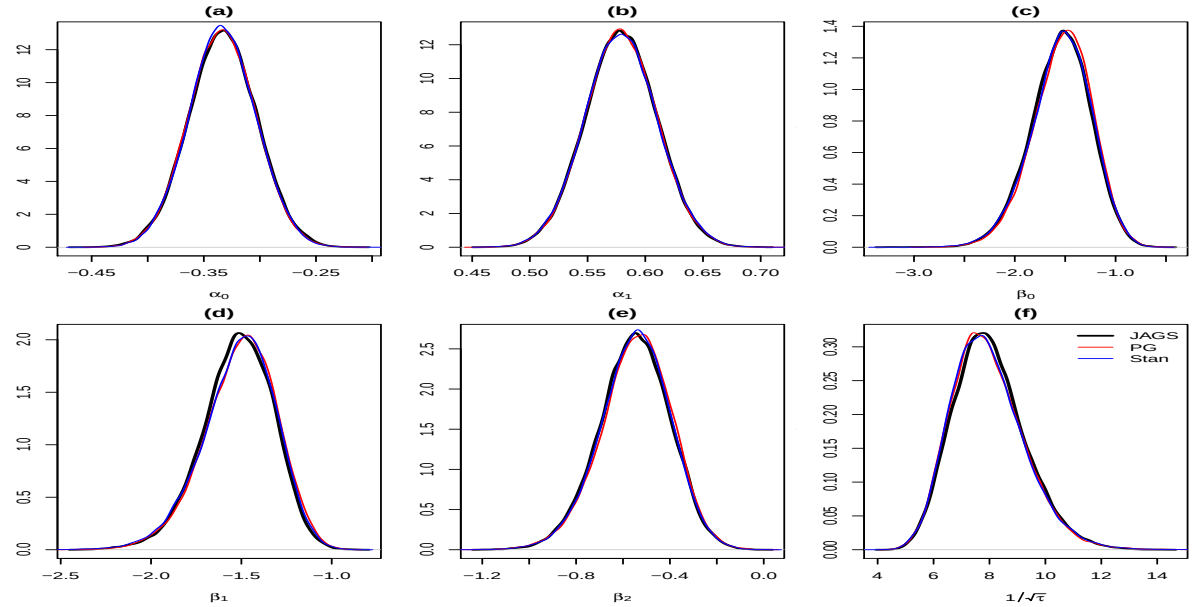

Figure 24. Posterior distributions of the parameters of the Bayesian spatial occupancy model using *JAGS*, *Stan* as well as the Pólya-Gamma formulation for the Cape weaver data set ((a) = α_0, (b) = α_1, (c)= β_0, (d)= β_1, (e) = β_2, (f) = $\frac{1}{\sqrt{\tau}}$).

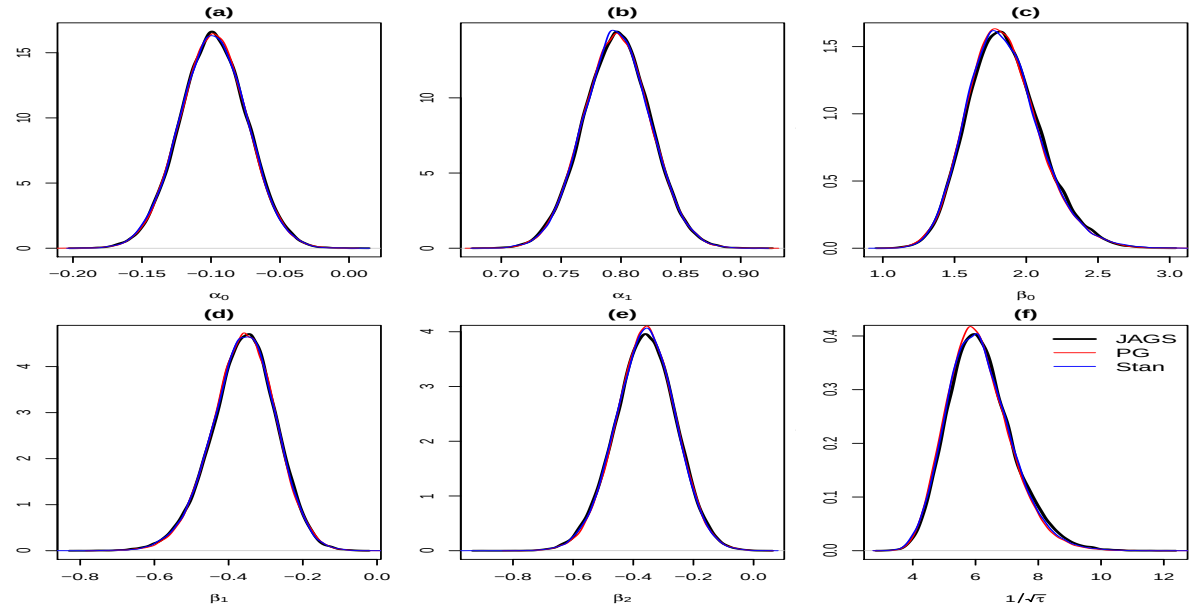

Figure 25. Posterior distributions of the parameters of the Bayesian spatial occupancy model using *JAGS*, *Stan* as well as the Pólya-Gamma formulation for the helmeted guineafowl data set ((a) = α_0, (b) = α_1, (c)= β_0, (d)= β_1, (e) = β_2, (f) = $\frac{1}{\sqrt{\tau}}$).

4.H Certain lagged sample autocorrelation functions.

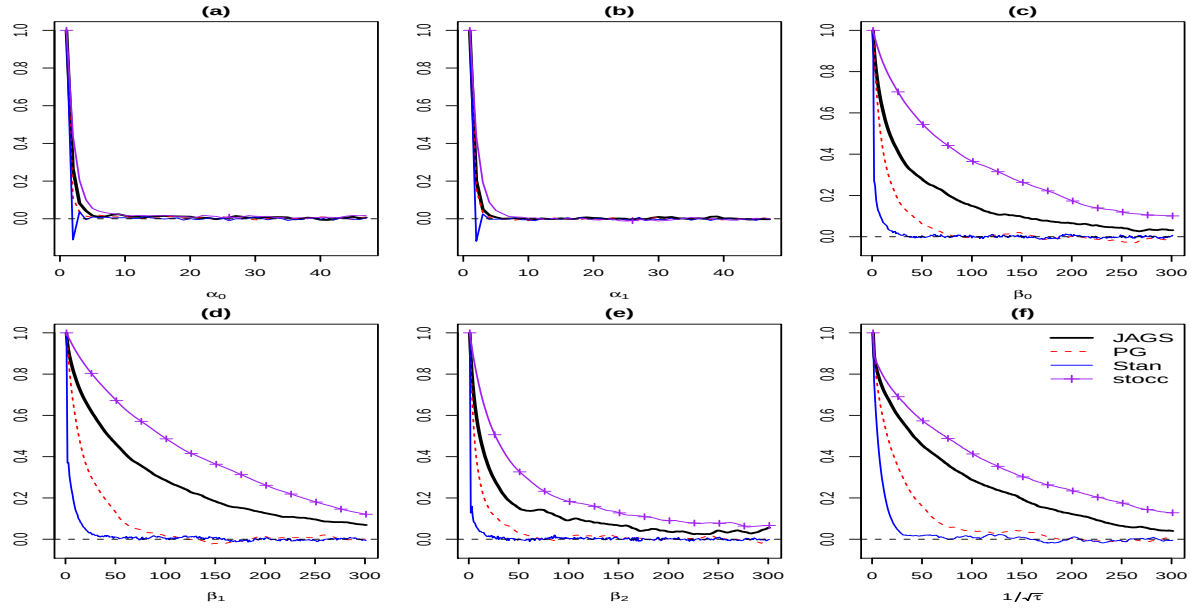

Figure 26. Estimated lagged sample autocorrelations of the posterior samples of the parameters of the Bayesian spatial occupancy model using *JAGS*, the Pólya-Gamma formulation, *Stan* and *stocc* for the Cape weaver data set ((a) = α_0, (b) = α_1, (c)= β_0, (d)= β_1, (e) = β_2, (f) = $\frac{1}{\sqrt{\tau}}$).

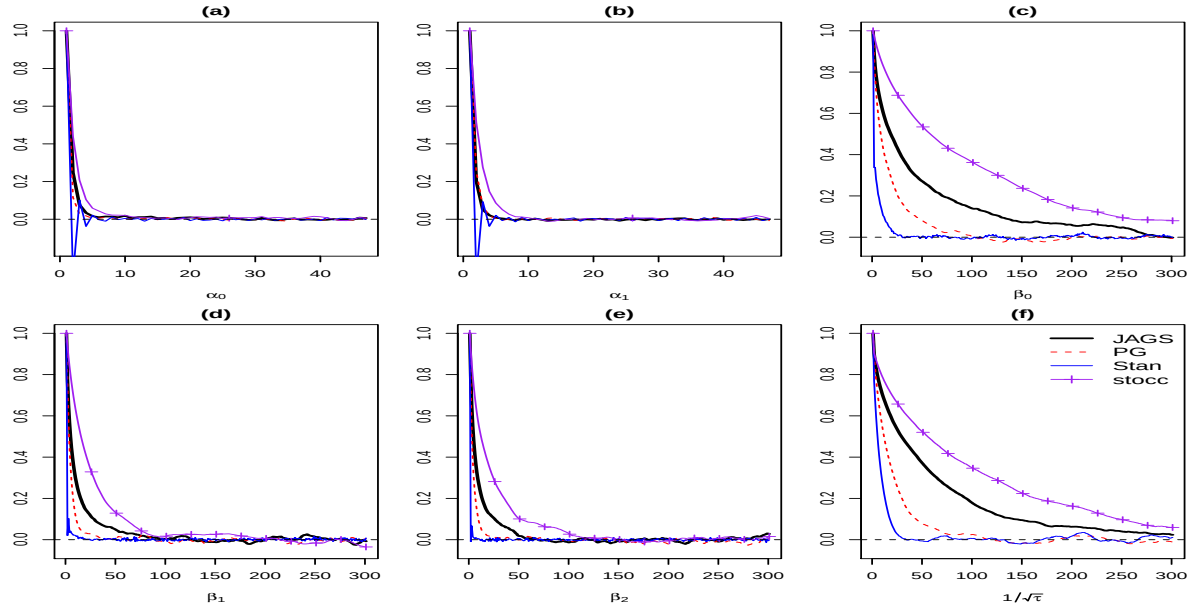

Figure 27. Estimated lagged sample autocorrelations of the posterior samples of the parameters of the Bayesian spatial occupancy model using *JAGS*, the Pólya-Gamma formulation, *Stan* and *stocc* for the helmeted guineafowl data set ((a) = α_0, (b) = α_1, (c)= β_0, (d)= β_1, (e) = β_2, (f) = $\frac{1}{\sqrt{\tau}}$).

4.I A worked example using simulated spatial data

Below we provide a worked example of how to fit the RSR model (in R) using *stocc*, *Stan* as well as *Rcppocc*. Ensure that you have installed the following R packages: *rstan*, *stocc* and *Rcppocc*. Installation instructions for *Rcppocc* can be found at `https://github.com/AllanClark/Rcppocc`.

We model the detection process using one covariate and the occupancy process using two covariates. In both cases, the intercept is included in the occupancy and detection processes. 188 spatial random effects are included in the spatial occupancy process. The same prior distributions as specified in the main part of the manuscript are used below.

Load the R data (contained in the *Rcppocc* package) file as well as some R packages.

```
require(stocc)
require(Rcppocc)

require(rstan)
rstan_options(auto_write = TRUE)

data(SpatSimData)
```

Create some R objects that will be used below.

```
xycords <- SimTable$xycords
psi <- SimTable$psi
z <- SimTable$z
alpha <- SimTable$alpha

Nvisits_surveyed <- SimTable$Nvisits_surveyed

Xmat <- SimTable$Xmat
Minv <- SimTable$Minv
Q <- SimTable$Q
Kmat <- SimTable$Kmat
Wmat <- SimTable$Wmat

pij <- SimTable$pij
Ysim <- SimTable$Ysim

nsites <- SimTable$nsites
Num_surveys <- SimTable$Num_surveys
N_maxvisits <- SimTable$N_maxvisits
```

The *stocc* fit

Fit the probit model using *stocc*.

```
#bundle data into the correct format for stocc
#occupancy dataframe at the subset of sites where surveys were performed
surveyIndex <- 1:Num_surveys #surveyIndex are ids of surveyed locations
```

```
siteIndex <- 1:nsites #ids for all sites considered

Site.Data <- as.data.frame( cbind( siteIndex, xycords, Xmat ) )
colnames(Site.Data)<- c("SiteName","Longitude","Latitude","PC1","PC2")

#detection dataframe - only at surveyed locations
PAdata <- na.omit(c(t(Ysim))) #presence-absence data
SiteNamereps <- rep(surveyIndex, Nvisits_surveyed) #the site names
nspp <- na.omit(c(t(Wmat))) #detection covariate
Longitudereps <- rep(xycords[surveyIndex,1], Nvisits_surveyed)
Latitudereps <- rep(xycords[surveyIndex,2], Nvisits_surveyed)

#at surveyed locations
Visit.data <- data.frame(SiteNamereps, Longitudereps, Latitudereps, nspp
    , PAdata)
colnames(Visit.data) <- c("SiteName", "Longitude", "Latitude", "nspp", "
    PAdata")

Names <- list(visit = list(site = "SiteName", obs = "PAdata"),
site = list(site = "SiteName", coords = c("Longitude", "Latitude")) )

Make.so.data <- make.so.data(visit.data = Visit.data,
site.data = Site.Data, names = Names)
```

Set prior distributions and some simulation settings.

```
#Set the prior distributions used
nalphas<-2
nbetas<-3

beta_m<-matrix( rep(0,nbetas), ncol=1)
sigma_inv_beta_p<-diag(nbetas)/1000 #prior inverse covariance for beta

alpha_m<-matrix( rep(0,nalphas), ncol=1)
sigma_inv_alpha_p<-diag(nalphas)/1000 #prior inverse covariance for
    alpha

numSpatre<-188

#mcmc settings
#These should be increased and are simply set to make it fairly fast to
    run!
nburnin <- 10000
niter <- 10000
nthin <- 1
```

The spatial.occupancy call.

```
spat_probit <- spatial.occupancy(detection.model= ~ nspp,
occupancy.model= ~ PC1 + PC2, spatial.model = list(model = "rsr",
threshold = 0.36, moran.cut = numSpatre, rho=1), so.data = Make.so.data,
prior = list(a.tau = 0.5, b.tau = 0.0005,
Q.b=sigma_inv_alpha_p, mu.b = alpha_m,
Q.g=sigma_inv_beta_p, mu.g = beta_m),
control = list(burnin = nburnin, iter = niter, thin = nthin))
```

The *Rcppocc* fit

```
spat_logit <- occSPATlogit(detection.model= ~ nspp, occupancy.model= ~
    PC1 + PC2,
spatial.model = list(threshold = 0.36, moran.cut = numSpatre, rho=1),
so.data = Make.so.data,
prior = list(a.tau = 0.5, b.tau = 0.0005,
tau=1,
Q.d=sigma_inv_alpha_p, mu.d = alpha_m,
Q.o=sigma_inv_beta_p, mu.o = beta_m),
control = list(ndraws =niter, percent_burn_in = 0.5))
```

The *Stan* fit

```
#see https://github.com/stan-dev/example-models/blob/master/misc/ecology
#/occupancy/occupancy.stan  - for simple occupancy code

#see http://mc-stan.org/users/documentation/case-studies/mbjoseph-
    CARStan.html
# - for sparse icar model

SSSO3 <- "
functions {
  //note the '188' has been hard coded.
  //this can be passed as an argument as well.

  real theta_lpdf(vector theta, real tau, matrix Minv){
  return 0.5*( 188*log(tau) - tau*quad_form(Minv, theta) );}
}

data {
   //note the '188' has been hard coded.

   int<lower=1> nsites;
   int N_maxvisits; //maximum number of survey visits
   int V[nsites]; //number of visits

   int<lower=0,upper=1> y[nsites, N_maxvisits]; //presence absence data
   matrix[nsites, 3] Xmat; // X variable
   matrix[nsites, 188] Kmat; // K variable
   real Wmat[nsites, N_maxvisits]; // W matrix - detection covariates
   matrix[188, 188] Minv; //K.t*Q*K
}

parameters {
   //note the '188' has been hard coded.

   vector[3] beta;  //occupancy slope params
   vector[2] alpha;  //detection slope params
   vector[188] theta;
   real<lower = 0> tau;
}

transformed parameters {
   vector[nsites] psi; // prob of of occurrence
   real pij[nsites, N_maxvisits]; // prob of detection
```

```
42
43    //calculate psi_i and pij
44    for (isite in 1:nsites){
45       //here Xmat[isite]*beta takes the 'isite'th row of Xmat and
46       //multiplies it by beta
47       psi[isite] = inv_logit( Xmat[isite]*beta + Kmat[isite]*theta  );
48
49       for (ivisit in 1:N_maxvisits){
50          pij[isite, ivisit] = inv_logit( alpha[1] +
51          Wmat[isite, ivisit]*alpha[2] );
52       }
53    }
54 }
55
56 model {
57    vector[nsites] log_psi;
58    vector[nsites] log1m_psi;
59
60    for (isite in 1:nsites) {
61       log_psi[isite] = log(psi[isite]);
62       log1m_psi[isite] = log1m(psi[isite]);
63    }
64
65    // priors
66    tau ~ gamma(0.5, 0.005);
67    theta ~ theta_lpdf(tau, Minv);
68    alpha ~ normal(0, 1000);
69    beta ~ normal(0, 1000);
70
71    // likelihood
72    for (isite in 1:nsites) {
73
74       if (sum(y[isite, 1:V[isite]]) > 0){
75          target += log_psi[isite] +
76          bernoulli_lpmf(y[isite, 1:V[isite]]|pij[isite,1:V[isite]]) ;
77       }else {
78          target += log_sum_exp(log_psi[isite] +
79          bernoulli_lpmf(y[isite, 1:V[isite]]|pij[isite, 1:V[isite]]),
80          log1m_psi[isite]);
81       }
82    }//end likelihood contribution
83 }
84 "
```

Compile the model and run it.

```
1 Model_code_covocc3 <- stan_model(model_code = SSS03)
2
3 V <- Nvisits_surveyed
4
5 #Note redefinition of some objects!
6
7 y <- Ysim
8 y[is.na(y)] <- 0 #replace all NA elements of y with a number. Stan does
     not allow NA's
9 Wmat[is.na(Wmat)] <- 0
10
11 Xmat <- cbind(1, Xmat[1:Num_surveys,]) #in the Stan code the intercept
```

```
   is included
Kmat <- Kmat[1:Num_surveys,]
nsites <- NROW(Xmat)

stanfit3 <- sampling(Model_code_covocc3, data=c("nsites","N_maxvisits","
   V","y",
"Xmat", "Kmat", "Wmat","Minv"), iter = niter, chains = 1, warmup =
   nburnin,
thin = 1, verbose=F)
```

4.J Data Accessibility:

All data files will be archived in Dryad [6]

[6] Refer to https://datadryad.org/stash/dataset/doi:10.5061/dryad.jt2002k for more details.

Chapter 5

Gibbs sampling algorithms for multi-species site occupancy models

Portions of the work in this chapter was presented at the following conferences: (i) a poster presentation at *34*th *International Workshop on Statistical Modelling, Guimarães, Portugal, 7-12 July 2019*; (ii) oral presentations at the *International Statistical Ecology conference*, St Andrews, Scotland, 2-6 July 2018 and the *South African Statistical Association conference*, Port Elizabeth, South Africa, 25-29 November 2019.

The R code used to undertake the analysis and simulation study can be found at `https://github.com/AllanClark/PhdAnalysisCode` while the code for the MSO package can be found at `https://github.com/AllanClark/MSO`.

Take note that the references for the article are all displayed at the end of the thesis.

'Allan Ernest Clark' wrote the paper and performed the mathematical development and analysis. 'Res Altwegg' and 'Allan Ernest Clark' reviewed the manuscript extensively. Both authors have read and approved the manuscript.

A Gibbs sampler for multi-species occupancy models

Allan E. Clark [1,2], Res Altwegg[1,2],

1. Department of Statistical Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, Cape Town, South Africa.
2. Centre for Statistics in Ecology, Environment and Conservation (SEEC), University of Cape Town, Cape Town, South Africa.

Abstract

Multi-species occupancy (MSO) models use detection-nondetection data from species observed at different locations to estimate the probability that a particular species occupies a particular geographical region. The models are particularly useful for estimating the occupancy probabilities associated with rare species since they are seldom observed when undertaking field surveys. In this paper, we develop Gibbs sampling algorithms that can be used to fit various Bayesian MSO models to detection-nondetection data. Bayesian analysis of these models can be undertaken using statistical packages such as *JAGS*, *Stan* and *NIMBLE*, however, since these packages were not developed specifically to fit occupancy models, one often experiences long run-times when undertaking analysis.

In a single season (single species) nonspatial and spatial occupancy modelling context, Clark and Altwegg (2019), show that special purpose Gibbs samplers can produce posterior chains that mix faster and have larger expected sampling rates (Holmes and Held, 2006) than those obtained using *JAGS* and *Stan*. These results suggest that such algorithms could potentially lead to significant reductions in the run-times of MSO models.

This paper illustrates how to fit MSO models when the detection and occupancy processes are modelled using logistic link functions and apply these methods to a camera-trapping study undertaken by Drouilly et al. (2018). Variable selection is undertaken using a reversible-jump Markov chain Monte Carlo (Barker and Link, 2013) algorithm. We found that the Gibbs sampling algorithm developed produces posterior samples that are identical to those obtained when using *Stan*, resulting in faster run-times and has a larger expected sampling rate than *Stan* when analysing the above-referenced data set.

Key words

Bayesian multi-species occupancy model, occupancy model, species richness, imperfect detection, reversible-jump Markov chain Monte Carlo.

5.1 Introduction

Species richness (defined as the number of distinct species in a region) is a measure of biodiversity (Dorazio et al., 2006) and is important from a conservation and management point of view since ecologists and nature conservationists use these estimates to prioritize conservation actions (Zipkin et al., 2009; Goijman et al., 2015). In most cases, species richness in an area is unknown and has to be estimated using observed data. Such data are collected by sampling a region a finite number of times. Some species might be easily observed while others might be more cryptic, thus making it more difficult to draw inferences about the community structure as a whole. Multi-species occupancy (MSO) models account for variable detection probabilities amongst species and are used to investigate the community structure of a group of species (Dorazio and Royle, 2005; Dorazio et al., 2006, 2011; Broms et al., 2016).

MSO models use detection-nondetection data from several species observed at different locations to estimate the probability that a species occupies a location. The models are particularly useful for estimating the occupancy probabilities associated with rare species (Devarajan et al., 2020). The observed detections of these species are usually low which results in imprecise occupancy probabilities if single species occupancy models are fit to such data. MSO models utilize random effects and allow the parameters associated with the regression effects of the model to be shared. This allows one to estimate detection and occupancy regression effects for species that would otherwise not be estimable (Tobler et al., 2015; Broms et al., 2016).

MSO models are mainly used to obtain estimates of species richness in a certain geographical region but can also be used to compare the biodiversity of different regions (Zipkin et al., 2010; Sauer et al., 2013). In certain geographical regions, species richness is known or well-understood (e.g., in well-studied areas) in which case these models are also used to uncover the biological drivers of biodiversity in a region (Zipkin et al., 2009; Ruiz-Gutiérrez et al., 2010; Drouilly et al., 2018).

Two formulations of the model have been developed in the literature. The first version assumes that **we know the species richness** in a certain geographical region (Broms et al., 2016) and it is assumed that all species of interest have been observed at least once. The second version of the model assumes that **species richness is an unknown random variable** where the aim is to estimate the number of unobserved species (Dorazio and Royle, 2005). Both model formulations are generally undertaken using Bayesian methods. They have extensively been used and computer code to fit these models are freely available online (Dorazio et al., 2006; Broms et al., 2016).

In the literature to date, the modelling process is mostly undertaken using statistical

packages such as *WinBUGS* (Lunn et al., 2000), *OpenBUGS* (Thomas et al., 2006), *JAGS* (Plummer, 2003), *NIMBLE* (de Valpine et al., 2017) and *Stan* (Carpenter et al., 2017). These packages allow the user to apply Bayesian methods to many different application areas and have not been developed specifically to undertake MSO models. *WinBUGS*, *OpenBUGS* and *JAGS* use a combination of Gibbs sampling (Geman and Geman, 1984), Metropolis Hastings (denoted as 'MH') steps (Metropolis et al., 1953; Hastings, 1970) as well as other custom Markov Chain Monte Carlo (MCMC) samplers (Brooks et al., 2011) to obtain posterior samples from the posterior distribution of the parameters of a Bayesian model. *NIMBLE* is flexible and allows the user to specify the sampling algorithms that should be used when undertaking posterior sampling. The above software packages with the exception of Stan, allow the user to sample from both discrete and continuous random variables. Stan in contrast only allows posterior sampling from continuous random variables which are undertaken using the *No-U-turn* Hamiltonian Monte Carlo sampler (Hoffman and Gelman, 2014).

Below we develop Gibbs sampling algorithms that can be used to fit MSO models. The Gibbs sampler is a MCMC method that is used to obtain samples from the posterior distribution of the parameters of a statistical model. The method can be used when all of the conditional posterior distributions of the parameters of a model are of a known form and sampling from the conditional posterior distributions are straightforward. The Gibbs sampler is known to exhibit slow mixing in certain cases although this shortcoming is often overcome by sampling groups of parameters in a block which improves mixing (Roberts and Sahu, 1997). In a single season (single species) nonspatial and spatial occupancy modelling context, Clark and Altwegg (2019), show that special purpose Gibbs samplers can produce posterior chains that mix faster and have larger expected sampling rates (Holmes and Held, 2006) than those obtained using JAGS and Stan. These results suggest that such algorithms could potentially lead to significant reductions in the run-times of MSO models.

Dorazio and Rodriguez (2012) and Johnson et al. (2013) developed Gibbs algorithms to obtain posterior samples for the parameters of a nonspatial and spatial single species occupancy model respectively. Both approaches assume that detection and occupancy processes are modelled using probit link functions, which enables the use of data augmentation (Tanner and Wong, 1987) to obtain closed form expressions of the conditional posterior distributions of the parameters of an occupancy model. Below we develop Gibbs sampling algorithms to undertake MSO models when species richness is known and thereafter we describe how the algorithm would change when species richness is unknown. Initially, we assume that the regression effects of the detection and occupancy processes are modelled using probit link functions and thereafter we extend the analysis to the use of logistic link functions. We do so since we observe that in certain model formulations, the use of

probit link functions does not lead to a simple Gibbs sampling algorithm when sampling from the posterior distribution of the parameters of an MSO model. In these cases, MH steps, which require the monitoring of various additional tuning parameters, could be used. Another reason for using the logistic link function is that it allows one to more easily interpret the regression effects of the model compared to the probit model. Take note that below we do not discuss joint species distribution models (Pollock et al., 2014; Tikhonov et al., 2020) and restrict the focus to occupancy modelling. A detailed review of MSO and its use can be found in Devarajan et al. (2020).

The paper commences by discussing the Bayesian formulation of an MSO model when species richness is assumed to be known in a geographical region while in Section 5.3 we develop a Gibbs sampler for the MSO model when species richness is unknown. In both cases, it is assumed that the detection and occupancy processes are modelled using probit link functions. In Section 5.4 we extend the previous work and consider the case when the detection and occupancy processes are modelled using logistic functions. We describe how the Gibbs algorithms are constructed and briefly comment how MSO models can be applied to different taxa or species groups. Before concluding, we re-analyse the data collected from a camera-trapping study undertaken by Drouilly et al. (2018) in order to highlight the use of the methods developed as well.

5.2 Known species richness - probit link function

The following formulation of the MSO model is similar to Broms et al. (2016)[1]. It is assumed that each site (J in total) can be surveyed numerous times potentially with unequal survey effort and that n_s species have been observed in total. The observed data is stored in a three-dimensional ragged-array $\boldsymbol{y}$ where $y_{i,j,k}$ represent the detection-nondetection of species i, at site j during survey k ($i = 1, \ldots, n_s;\ j = 1, \ldots, J;\ k = 1, \ldots, K_j$). The random variable $y_{i,j,k} = 1$ if species i is observed at site j during visit k and 0 otherwise. The partially observed occupancy states are stored in a matrix $\boldsymbol{z}$ where $z_{i,j} = 1$ if species i occupies site j while $z_{i,j} = 0$ if species i does not occupy site j. We assume that the occupancy process can be modelled using site specific covariates ($\boldsymbol{X}$) while the detection process can be modelled using survey-specific covariates ($\boldsymbol{V}$) where it is assumed that these covariates do not vary by species.

The model can now be specified using the following Bayesian hierarchical model,

$$z_{i,j}|\psi_{i,j} \sim \text{Bernoulli}(\psi_{i,j}),\ i = 1, \ldots, n_s;\ \ j = 1, \ldots, J,$$

[1] n_s = the number of observed species; J = the number of sites; K_j = the number of surveys undertaken at site j; $\sum_j = K_j$ = the total surveys undertaken; n_d = the number of detection covariates (including an intercept); n_o = the number of occupancy covariates (including an intercept); $\boldsymbol{X}$ is a $J \times n_o$ matrix; $\boldsymbol{z}$ is a $n_s \times J$ matrix; $\boldsymbol{V}$ is a $\left(\sum_{j=1}^{J} K_j \times J\right) \times n_d$ matrix.

$$y_{i,j,k}|z_{i,j}, p_{i,j} \sim \text{Bernoulli}(z_{i,j}p_{i,j,k}), k = 1, \ldots, K_j$$

where occupancy and detection probabilities are modelled using probit link functions, $\psi_{i,j} = \Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)$ and $p_{i,j,k} = \Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i)$ $\forall(i,j,k)$. The Bayesian formulation is completed by specifying the prior distributions used. The regression effects of the detection and occupancy processes are modelled using multivariate Gaussian distributions and we specifically assume that

$$\begin{aligned}\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}) \text{ with } \boldsymbol{\mu}_{\boldsymbol{\alpha}} \sim \mathcal{N}(\boldsymbol{0}, a^2\boldsymbol{I}_{n_d}) \text{ and}\\ \boldsymbol{\beta}_i|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o}) \text{ with } \boldsymbol{\mu}_{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{0}, b^2\boldsymbol{I}_{n_o}).\end{aligned}$$

The coefficient vectors $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ represent the community-level effects associated with the detection and occupancy processes respectively and are each assigned a common Gaussian hyper-prior distribution. By doing so, one allows the regression effects to *share information* since often the detection and occupancy probabilities are not estimable for rare animal species due to lack of detections (Tobler et al., 2015).

Finally, $\sigma_{\boldsymbol{\alpha}}$ and $\sigma_{\boldsymbol{\beta}}$ are modelled using a half-Cauchy distribution (Gelman, 2006) and can be specified hierarchically as

$$\sigma^2_{\boldsymbol{\alpha}}|\nabla_{\boldsymbol{\alpha}} \sim \text{IG}\left(0.5, \frac{1}{\nabla_{\boldsymbol{\alpha}}}\right), \nabla_{\boldsymbol{\alpha}} \sim \text{IG}\left(0.5, \frac{1}{A^2}\right) \text{ and} \tag{5.1}$$

$$\sigma^2_{\boldsymbol{\beta}}|\nabla_{\boldsymbol{\beta}} \sim \text{IG}\left(0.5, \frac{1}{\nabla_{\boldsymbol{\beta}}}\right), \nabla_{\boldsymbol{\beta}} \sim \text{IG}\left(0.5, \frac{1}{B^2}\right) \tag{5.2}$$

where A and B are known hyper-parameters.

Posterior samples can be obtained from the above MSO model by using a MH algorithm although we instead derive a Gibbs algorithm that uses the data augmentation scheme developed by Albert and Chib (1993) to represent a probit link function by means of latent Gaussian random variables. The full conditional distributions needed to undertake a Gibbs sampler are all of a familiar form and are shown in Appendix 5.A.1. In Appendix 5.A.2 we show how similar conditional posterior distributions can be obtained if site-specific covariates are available but no survey-specific covariates are available for the detection process. In this case, site-specific covariates are used as proxies for the survey-specific covariates.

5.3 Unknown species richness - probit link function.

Dorazio and Royle (2005) developed a model that could be used to estimate the species richness (denoted as N) in a geographical region. They assume that there exists a super-population of species (denoted as S) that consists of the observed species, n_s, as well as

additional unseen species, $S - n_s$. Recall that in Section 5.2 survey occasion was explicitly included in the model formulation. Here all information relating to survey occasions are grouped together such that the observed data $\boldsymbol{y}_{n_s}$ consists of a $n_s \times J$ matrix of the total number of detections for each species at each location (from K_j surveys, $j = 1, \ldots, J$). Since N is unknown, an $(S - n_s) \times J$ matrix of zeros are introduced which represents the data associated with the unobserved species.

A latent indicator variable w_i is introduced which takes on the value 1 if species i in the super-population occurs in the geographical region under investigation and 0 otherwise. From the above discussion, $w_i = 1$ for $i = 1, \ldots, n_s$ and the species richness in a geographical region is $N = \sum_{i=1}^{S} w_i$. The latent variable $z_{i,j}$ takes on the value 1 if species i occupies location j and 0 otherwise. The w_i indicator is modelled using a Bernoulli distribution with success parameter Ω while $z_{i,j}|w_i$ is modelled as a Bernoulli random variable with success probability, $w_i\psi_{i,j}$. The detection process (conditional on $w_i = 1, z_{i,j} = 1$) is modelled using a binomial distribution such that

$$p(y_{i,j}|z_{i,j} = 1, p_{i,j}, w_i = 1) = \binom{K_j}{y_{i,j}} p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{K_j - y_{i,j}}. \tag{5.3}$$

Using the above description it is clear that the model can be formulated hierarchically as

$$\begin{aligned}
w_i &\sim \text{Bernoulli}(\Omega),\ \forall i = 1, \ldots, S \\
z_{i,j}|w_i, \psi_{i,j} &\sim \text{Bernoulli}(w_i\psi_{i,j}),\ \forall i = 1, \ldots, S;\ \ \forall j = 1, \ldots, J, \\
y_{i,j}|z_{i,j}, w_i, p_{i,j} &\sim \text{Binomial}(K_j,\ w_i z_{i,j} p_{i,j}), \\
p_{i,j} &= \Phi(\boldsymbol{v}_j^T \boldsymbol{\alpha}_i) \text{ with } \boldsymbol{\alpha}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma_{\boldsymbol{\alpha}}^2 \boldsymbol{I}_{n_d}), \\
\psi_{i,j} &= \Phi(\boldsymbol{x}_j^T \boldsymbol{\beta}_i) \text{ with } \boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2 \boldsymbol{I}_{n_o}),
\end{aligned}$$

where $\boldsymbol{v}_j$ and $\boldsymbol{x}_j$ variables are site-specific covariates used to model the detection and occupancy processes respectively.

Since the detection process is a binomial distribution, the use of data augmentation method developed previously to sample from the detection regression effects is more cumbersome to use since the number of latent variables required to perform the data augmentation step increases from 1 to K_j for each i. The remaining parameters however have simple conditional posterior distributions if we use data augmentation to obtain the posterior distribution of the occupancy regression coefficients. A derivation of the resulting conditional posterior distributions can be found in Appendix 5.B.

5.4 The use of logistic link functions in multi-species occupancy models.

In Appendix 5.C and 5.D we show that if logistic link functions are used to model the regression effects of the detection and occupancy processes, a Gibbs sampling algorithm, with closed form conditional posterior distributions for **all parameters** of the MSO models in Sections 5.2 and 5.3, can be derived. A Gibbs sampler is preferable to the use of a MH algorithm since the former does not require any 'tuning' when undertaking the MCMC sampling.

As an example, consider the MSO model developed in Section 5.3 but assume that the regression effects for detection and occupancy are modelled as

$$\begin{aligned}\text{logit}(p_{i,j}) &= \boldsymbol{v}_j^T\boldsymbol{\alpha}_i \text{ where } \boldsymbol{\alpha}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}),\\ \text{logit}(\psi_{i,j}) &= \boldsymbol{x}_j^T\boldsymbol{\beta}_i \text{ where } \boldsymbol{\beta}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o})\end{aligned}$$

and both $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ are modelled hierarchically as done previously.

The conditional posterior distributions of most of the parameters remain the same as those defined in Section 5.3 although the conditional posterior distributions of $\boldsymbol{\alpha}_i|.$ and $\boldsymbol{\beta}_i|.$ are slightly amended as detailed below.

When $w_i = 0$, $\boldsymbol{\beta}_i|. \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o})$ while when $w_i = 1$, posterior samples of $\boldsymbol{\beta}_i$ can be obtained by sampling from a Pólya-Gamma distribution and Gaussian distribution in turn. Specifically,

$$\begin{aligned}\boldsymbol{\omega}_{i,j}^{(\beta)}|\boldsymbol{\beta}, w_i = 1, . &\sim \text{PG}(1, \boldsymbol{x}_j^T\boldsymbol{\beta}_i),\ \forall j = 1, \ldots, J \text{ and}\\ \boldsymbol{\beta}_i|\boldsymbol{y}, \boldsymbol{\omega}_i^{(\beta)}, w_i = 1, . &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i})\end{aligned}$$

where the 'PG' notation denotes a Pólya-Gamma random variable, $\boldsymbol{\omega}_i^{(\beta)} = \left[\omega_{i,1}^{(\beta)}, \ldots, \omega_{i,J}^{(\beta)}\right]^T$, $\tau_{\boldsymbol{\beta}} = \frac{1}{\sigma^2_{\boldsymbol{\beta}}}$ and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\beta}_i} &= \left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{D}_{\boldsymbol{\beta}}\boldsymbol{X}\right)^{-1}\\ \boldsymbol{\mu}_{\boldsymbol{\beta}_i} &= \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i}\left(\boldsymbol{X}^T\boldsymbol{\Theta}_{\boldsymbol{\beta}} + \tau_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}}\right).\end{aligned}$$

The matrices $\boldsymbol{D}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\beta}}$ are defined as, $\boldsymbol{D}_{\boldsymbol{\beta}} = \text{diagonal}\left[\omega_{i,1}^{(\beta)}, \ldots, \omega_{i,J}^{(\beta)}\right]$ and $\boldsymbol{\Theta}_{\boldsymbol{\beta}} = [z_{i,1} - 1/2, \ldots, z_{i,J} - 1/2]^T$.

The conditional posterior distribution of $\boldsymbol{\alpha}_i|.$ can be obtained similarly such that

$$\begin{aligned}\boldsymbol{\omega}_{i,j}^{(\alpha)}|\boldsymbol{\alpha}, w_i = 1, . &\sim \text{PG}(K_j, \boldsymbol{x}_j^T\boldsymbol{\alpha}_i),\ \forall j = 1, \ldots, j^*\\ \boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\omega}_i^{(\alpha)}, w_i = 1, . &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i})\end{aligned}$$

where $\boldsymbol{\omega}_i^{(\boldsymbol{\alpha})} = \left[\omega_{i,1}^{(\boldsymbol{\alpha})}, \ldots, \omega_{i,j^*}^{(\boldsymbol{\alpha})}\right]^T$, $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma_{\boldsymbol{\alpha}}^2}$ and

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i} &= \left(\tau_{\boldsymbol{\alpha}} \boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T \boldsymbol{D}_{\boldsymbol{\alpha}} \tilde{\boldsymbol{V}}\right)^{-1} \\ \boldsymbol{\mu}_{\boldsymbol{\alpha}_i} &= \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i} \left(\tilde{\boldsymbol{V}}^T \boldsymbol{\Theta}_{\boldsymbol{\alpha}} + \tau_{\boldsymbol{\alpha}} \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right).\end{aligned}$$

Define $\tilde{\boldsymbol{V}}$ as the elements of $\boldsymbol{V}$ associated with $z_{i,j} = 1$ while the matrices $\boldsymbol{D}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\alpha}}$ are defined as, $\boldsymbol{D}_{\boldsymbol{\alpha}} = \text{diagonal}\left[\omega_{i,1}^{(\boldsymbol{\alpha})}, \ldots, \omega_{i,j^*}^{(\boldsymbol{\alpha})}\right]$ and $\boldsymbol{\Theta}_{\boldsymbol{\alpha}} = [y_{i,1} - K_1/2, \ldots, y_{i,j^*} - K_{j^*}/2]^T$ for an appropriate value of j^*.

Complete derivations of the conditional posterior distributions of the MSO model can be found in Appendix 5.C while Appendix 5.D contains the required conditional distributions when species richness is assumed known. In the above formulations of the MSO model, it is assumed that both the detection and occupancy regression effects are modelled as realisations from a common Gaussian distribution. These models can easily be extended by adding one additional level of complexity and assume that the regression effects of different groups (e.g., bird guilds, taxa or species groups) are modelled as realisations from different Gaussian distributions. The required algorithm is included in Appendix 5.D.2.

5.5 Application

As an application of the methods developed, we re-analyse the data collected from a camera-trapping study undertaken by Drouilly et al. (2018). The Anysberg Nature Reserve (33° 31′ S, 20° 37′ E) in South Africa (referred to as *Anysberg*) is the study area. The camera trapping design consisted of 156 camera trap sites (deployed from the end of September 2013-May 2014) each placed at a location within a 2 km grid design (Kinnaird and O'Brien, 2012). Data were collected using Bushnell Trophy CAM HG (Bushnell Outdoor Products, Overland Park, Kansas) camera traps and consisted of the processing of camera images obtained when cameras were triggered in the field. The study specifically focused on terrestrial vertebrates > 0.5 kilograms in mass. A detailed explanation of camera trap survey design can be found in Section 2.2 of Drouilly et al. (2018). A list of the species observed at least once at Anysberg (35 in total) during the course of the study can be found in Appendix 5.E.

Since the number of detections obtained daily were low, it was decided to define a sampling occasion as a 6 day period. From the captured images an observation matrix was constructed which consisted of the observed number of times species i (for $i = 1, \ldots, n_s = 35$) was detected at site j ($j = 1, \ldots, J = 156$) during K_j sampling occasions. The observed data is stored as $[y_{i,j}]$.

Below we use the observed species at Anysberg and relate the occupancy and detection

processes to explanatory variables. We assume that **species richness is known** and fit the MSO model described in Section 5.2 (assuming logistic link functions) to the data. The occupancy process is modelled as a function of *elevation* (measured as metres above sea-level), the *modified soil-adjusted index* (*MSAVI2*) (Qi et al., 1994) and a *PreyIndex*[2] MSAVI2 is a vegetation index similar to the normalised difference vegetation index. *MSAVI2* was converted into a binary indicator variable (named *MSAVI2ind*) such that *MSAVI2ind* equals 1 if the standardised variable *MSAVI2* is positive. The detection process is modelled as a function of a binary indicator variable, *trail* and a habitat factor variable. The *trail* variable equals 1 when a camera is placed on a trail and 0 otherwise. The habitat factor variable has three levels namely, *plain*, *river* and *mountain* where the *mountain* level is used as the baseline.

We fit all model combinations (32 in total, see Appendix 5.F) and assume that all models include an intercept in the detection and occupancy process. We further assume that we are in a $\mathcal{M}$-complete setting (Bernardo and Smith, 1994) and acknowledge that other variables might well be the true drivers of both the occupancy and detection process but choose to use the ones listed above. All MCMC sampling was undertaken using the authors' code. For all models, three chains of 100 000 draws were used with a burn-in proportion of one-third. The following hyper-parameters were used: $a^2 = 2.25^2 = b^2$ and $A^2 = 2.25^2 = B^2$. MSO models produce significant amounts of outputs and are computationally expensive. Because of this, we thinned the resulting MCMC chains using a factor of ten. Various convergence tests[3] were undertaken to assess that the MCMC chains had converged (Smith, 2007).

Exploratory (with-in sample) model selection was initially undertaken by first considering the WAIC (Watanabe, 2010) and the approximate leave-one-out cross-validation scores (LOO) using Pareto smoothed importance sampling (Vehtari et al., 2017) for each of the candidate models. In order to assess the variability of the model selection statistics ten independent MCMC runs were undertaken and two goodness-of-fit statistics namely the Bayesian p-value (Meng et al., 1994) and the sampled Bayesian p-value (Zhang, 2014) were also calculated. Gosselin (2011) show that asymptotically the distribution of the sampled Bayesian p-value is uniformly distributed over the interval $[0, 1]$ which suggests that misspecified models will not have uniformly distributed sample Bayesian p-values. The Bayesian posterior probability of each of the models were calculated using the version of the Reversible jump MCMC (RJMCMC) algorithm developed by Barker and Link

[2]The *PreyIndex* variable was constructed as the number of independent pictures for certain types of prey species considered in the study divided by the number of camera trap nights across the whole survey multiplied by 100. Pictures were assumed to be independent if they were taken more than 30 minutes apart.

[3]Trace plots were examined were examined to informally assess the mixing of the three chains while Geweke and Brooks-Gelman-Rubin tests were performed to assess the burn-in sample size and assess convergence of the Markov chains respectively.

(2013) which were then used to undertake Bayesian model averaging (Hoeting et al., 1999). We assume a-priori that each model is equally likely. Refer to Appendix 5.G for more details regarding the above mentioned model selection methods.

5.5.1 Application Results

Both model selection measures require the calculation of the variance of the logarithm of the integrated likelihood evaluated using the posterior samples at each MCMC iteration. Based on ten independent MCMC runs, we found that both model selection measures were not accurately estimated since the previous mentioned variance calculation was often greater than 0.4 thus rendering both model selection methods unreliable (Vehtari et al., 2017).

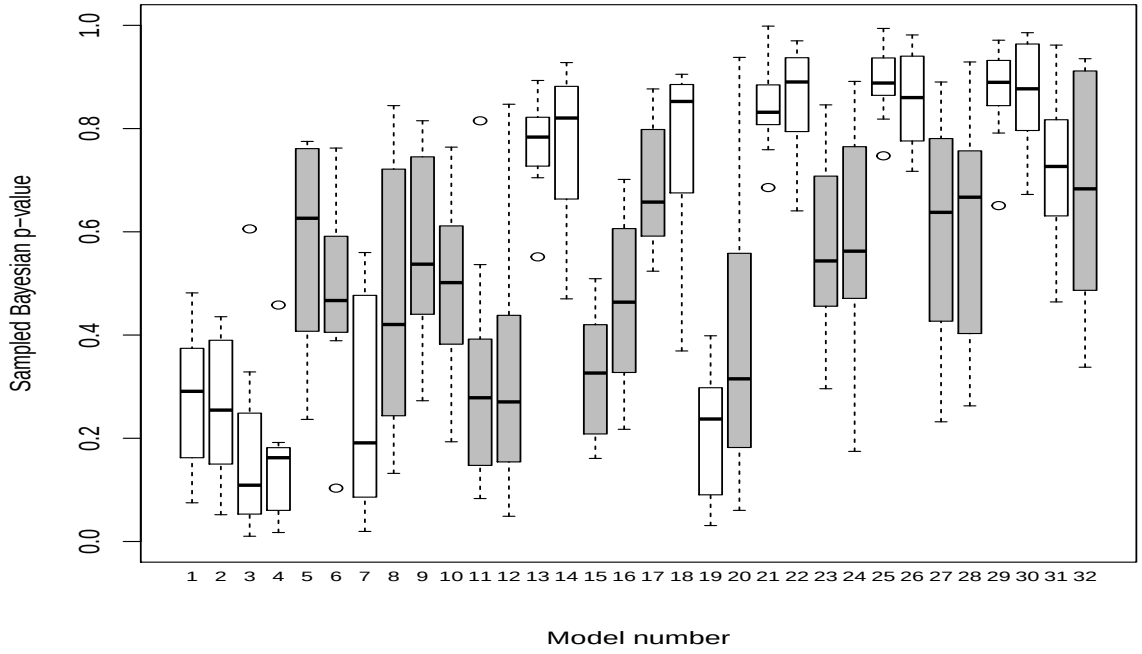

Figure 28. The sampled Bayesian p-values of each model based on 10 independent replications. The highlighted boxplots indicate that the assumption that the distribution of the sampled Bayesian p-values are uniformly distributed over the interval $[0, 1]$ cannot be rejected for these models when testing at a 30% significance level. The model associated with the 'Model number' can be found in Appendix 5.F.

The variability of the Bayesian p-values were of an order of magnitude smaller than those observed for the sampled Bayesian p-values (see Appendix 5.H). Anderson Darling tests (Anderson and Darling, 1954) using the sampled Bayesian p-values obtained from ten independent MCMC runs, indicate that the assumption that the distribution of the sampled Bayesian p-values are uniformly distributed over the interval $[0, 1]$ cannot be

rejected for fifteen of the thirty two models (assuming a 30% significance level[4]). These models have been highlighted in grey in Figure 28, which displays the boxplots of the sampled Bayesian p-values based on ten independent MCMC runs. The models and the p-values of the Anderson-Darling tests are tabulated in Table 16. Given the small sample size (of ten), these tests tentatively suggests that models that only include an intercept term in the occupancy process are misspecified. These results also suggest that none of the variables considered always lead to misspecified models and thus all variables do appear to have some predictive ability of occupancy for the species considered in this data set.

Table 16. The list of models associated with non-significant Anderson-Darling tests when testing the null hypothesis that the distribution of the sampled Bayesian p-values are uniformly distributed over the range $[0, 1]$; the respective p-value of the test as well as the median sampled Bayesian p-value of the associated models.

Model	**Occupancy covariates**	**Detection covariates**	**p-value**	**Median sampled Bayesian p-value**
5	elevation	1	0.67	0.63
6	elevation	trail	0.78	0.47
8	elevation	trail + plain + river	0.76	0.42
9	MSAVI2ind	1	0.74	0.54
10	MSAVI2ind	trail	0.82	0.5
11	MSAVI2ind	plain + river	0.30	0.28
12	MSAVI2ind	trail + plain + river	0.31	0.27
15	elevation + MSAVI2ind	plain + river	0.42	0.33
16	elevation + MSAVI2ind	trail + plain + river	0.80	0.46
17	PreyIndex	1	0.43	0.66
20	PreyIndex	trail + plain + river	0.50	0.32
23	elevation + PreyIndex	plain + river	0.76	0.54
24	elevation + PreyIndex	trail + plain + river	0.64	0.56
27	MSAVI2ind + PreyIndex	plain + river	0.62	0.61
28	MSAVI2ind + PreyIndex	trail + plain + river	0.58	0.67
32	elevation + MSAVI2ind + PreyIndex	trail + plain + river	0.31	0.68

Conn et al. (2018) suggests that the Bayesian p-values should not be used to undertake model selection, but should rather be used to identify models that display clear signs of model misspecification. In our context, models 3, 4, 21, 25, 26, 29, and 30 appear to be misspecified since they either have a Bayesian p-value above 0.8 or are smaller than 0.2. These models either only include an intercept term in the detection or occupancy process; do not include the *elevation* variable in the occupancy process or only includes the *trail* variable in the detection process.

The resulting posterior model probabilities from the RJMCMC algorithm are displayed

[4]A liberal significance level was selected since the sample size is very small.

Table 17. The posterior model probabilities for the five best models obtained using RJMCMC

Model	Occupancy covariates	Detection covariates	Posterior probability	Posterior odds
6	elevation	trail	0.8543	1
8	elevation	trail + plain + river	0.1442	5.92
22	elevation + PreyIndex	trail	0.0011	776.64
24	elevation + PreyIndex	trail + plain + river	0.0002	4 271.5
5	elevation	1	0.0001	8 543

in Table 17. The models listed contain a combination of the *elevation* and *PreyIndex* variable in the occupancy process as well as a combination of the indicator variables *trail*, *plain* and *river* in the detection process. From these results, we observed that Model 6 has the largest posterior probability with a value of 0.85, Model 8 has the second largest posterior probability with a value of 0.14 while all other models have very low posterior mass associated with them.

From an examination of the population regression effects in Table 22 in Appendix 5.I it can be seen that the mean population regression effects for plain and river in the detection process (of Models 8 and 24) both have 95% credibility intervals that contain zero. Similarly, the PreyIndex variable in Models 22 and 24 also have 95% credibility intervals that contain zero. Table 18 contains the Bayesian model averaged population regression effects[5] as well as the posterior effect probabilities (the probability that a population regression effect is non-zero). The results clearly indicate that trail and elevation have non-zero regression effects in the detection and occupancy processes respectively while the remaining regression effects have a high probability of being zero. These results indicate that on average one is more likely to detect one of the animal species investigated off a trail than on it and that on average no significant differences in detectability of animal species are observed between the different habitat types. We found that the mean population regression effect for elevation is negative indicating that on average higher occupancy probabilities are observed at lower elevation levels than at higher ones.

Figure 29 displays caterpillar plots highlighting the mean community regression effects as well as the 95% credibility intervals for the species specific regression effects of *trail* (a) and *elevation* (b) for the detection and occupancy processes respectively. Of the 35 species photographed, the occupancy probability was strongly related to *elevation* for 19 species. The results indicate that Smith's rock hare (*Pronolagus rupestris*), klipspringers

[5]From Raftery (1993), $\hat{\triangle}_k = \mathbb{E}(\triangle_k|\mathcal{M}_k, \text{Data})$ such that $\mathbb{E}(\triangle|\text{Data}) = \sum_k \mathbb{E}(\triangle_k|\mathcal{M}_k)\Pr(\mathcal{M}_k|\text{Data})$ and $\text{var}(\triangle|\text{Data}) = \left(\sum_k \left(\text{var}(\triangle|\mathcal{M}_k, \text{Data}) - \hat{\triangle}_k^2\right)\Pr(\mathcal{M}_k|\text{Data})\right) - \mathbb{E}(\triangle_k|\mathcal{M}_k, \text{Data})^2$. The posterior effect probabilities are obtained by summing the posterior model probabilities across models for each variable (Hoeting et al., 1999).

Table 18. The Bayesian model averaged community-level regression effects. (Posterior mean, Standard deviation as well as posterior effect probabilities.) Regression effects cannot be calculated if $\Pr(\beta \neq 0|\text{Data})$ and is captured as '-'.

Coefficient	**Mean**	**Std**	$\Pr(\boldsymbol{\beta} \neq 0\vert \text{ Data})$
$\boldsymbol{\alpha}_{\text{trail}}$	-0.54	0.21	1
$\boldsymbol{\alpha}_{\text{plain}}$	0.02	0.10	0.14
$\boldsymbol{\alpha}_{\text{river}}$	0.02	0.11	0.14
$\boldsymbol{\alpha}_{\text{MSAVI2ind}}$	-	-	0
$\boldsymbol{\beta}_{\text{elevation}}$	-0.59	0.22	1
$\boldsymbol{\beta}_{\text{PreyIndex}}$	0.0003	0.01	0.001
$\boldsymbol{\beta}_{\text{MSAVI2ind}}$	-	-	0

(*Oreotragus oreotragus*) and leopard's (*Panthera pardus*) were more likely to occupy high lying regions of the reserve compared to many of the antelope species that preferred low-lying regions of the reserve. In general, the detection of most carnivore species were negatively related to *trail* while the detection of species such as the hyrax (*Procavia capensis*) and the duiker (*Sylvicapra grimmia*) were positively related to *trail.*

We found that the above results were not sensitive to the specification of the prior distribution. To this end, in Appendix 5.K we display statistics pertaining to the community-level regression effects for Model 6 when run using three chains of 100 000 in length. A burn-in sample of one-third is specified thereafter the resulting chains were thinned by retaining every 10^{th} observation. The different prior distributions were specified as follows: Prior 1: $a^2 = 2.25^2 = b^2$, $A^2 = 2.25^2 = B^2$; Prior 2: $a^2 = 2.25^2 = b^2$, $A^2 = 2.25 = B^2$; Prior 3: $a^2 = 5^2 = b^2$, $A^2 = 5^2 = B^2$; Prior 4: $a^2 = 5^2 = b^2$, $A^2 = 5 = B^2$.

5.6 Discussion and Conclusions

Multi-species occupancy models are important tools for statistical ecologists since they allow one to undertake community-level as well as species-specific inference by using detection-nondetection data of many species at different locations (Dorazio and Royle, 2005; Dorazio et al., 2006; Broms et al., 2016). The information across species are pooled together which allows one to obtain occupancy probabilities of rare species which would otherwise not be possible due to the low detectability of these species.

Here, we develop Gibbs sampling algorithms when probit or logistic link functions are used to model the detection and occupancy processes respectively and show how this can be done for a number of different versions of the MSO models. We also specifically show that the algorithms for MSO models (that use a logistic link function) are similar to those developed for single season occupancy models as well as single season spatial occupancy models where all algorithms use the work of Polson et al. (2013) and Clark and Altwegg

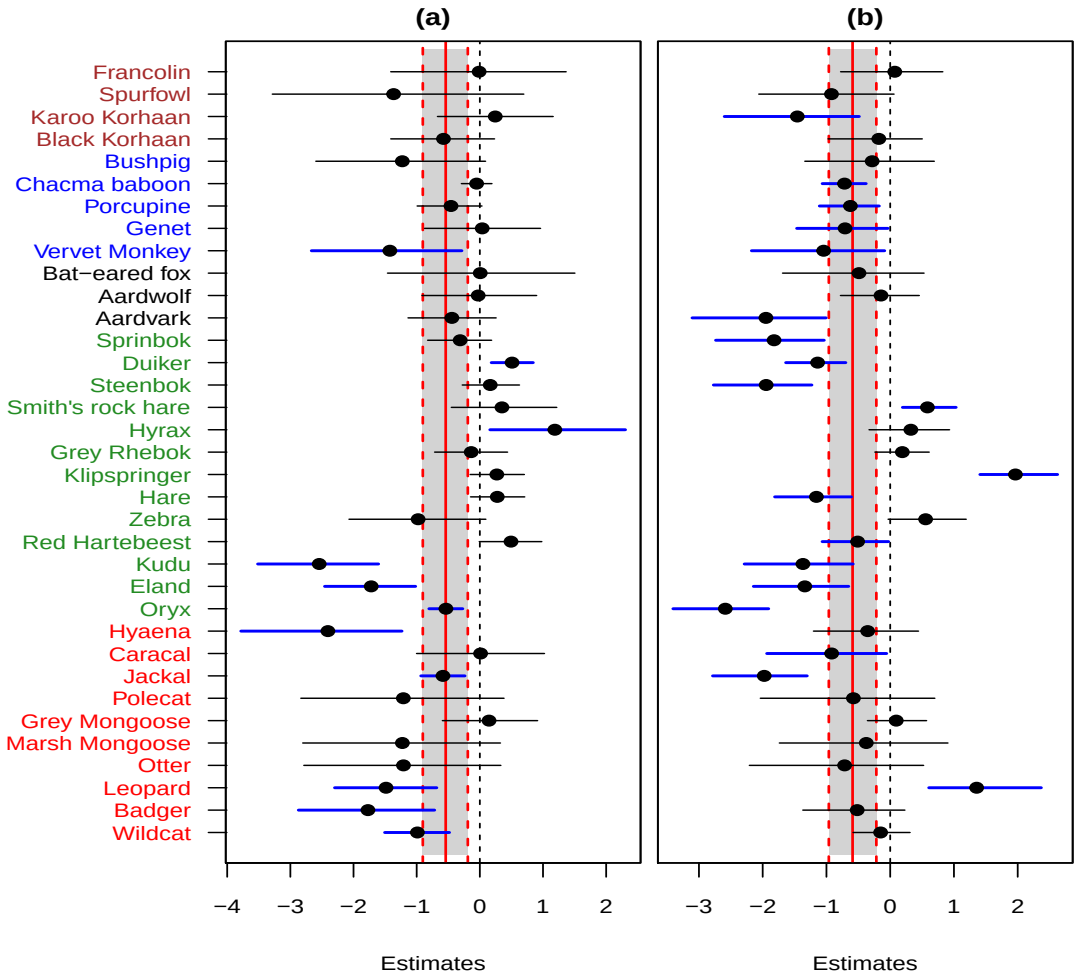

Figure 29. Caterpillar plots highlighting the mean community regression effects as well as the 95% credibility intervals for the species specific regression effects of *trail* (a) and *elevation* (b). Credibility intervals in bold do not overlap 0. The thick dashed lines indicate the 95% equal tail credibility interval for the mean community response to each variable. All quantities were calculated using the mixture distributions obtained when undertaking Bayesian model averaging. The results were grouped into the five groups namely carnivores (red), herbivores (green), insectivores (black), omnivores (blue) and birds (brown).

(2019) to develop efficient Gibbs sampling algorithms.

In our application we found that model selection methods such as WAIC (Watanabe, 2010) and the approximate leave-one-out cross-validation scores using Pareto smoothed importance sampling (Vehtari et al., 2017) were unreliable for the data considered. A possible reason for this being the sparsity of data at a number of sites. Model selection was undertaken using Reversible jump MCMC (Barker and Link, 2013) while the occupancy surfaces for all species considered were obtained using Bayesian model averaging (see Appendix 5.J).

A single MSO model took approximately $12-13$ minutes to fit compared to approximately 4 hours when using *Stan*. All calculations were performed using R 4.0.4 (R Core Team, 2014) on a Windows 10 desktop with a AMD Ryzen 7 3700X 8-Core Processor, a clock speed of 3.59 GHz and 64 GB of RAM. On the same system, the RJMCMC algorithm took approximately one hour to complete. These run-times were **not found to be prohibitively long** and thus we would encourage others to consider using the technique when undertaking model selection. We wrote our own code to undertake the RJMCMC algorithm although the R package *rjmcmc* (Gelling et al., 2019) could also be used.

Appendix

5.A Derivation of the conditional distributions required to undertake a Gibbs sampler for a MSO model when species richness is known.

5.A.1 Bernoulli detection process (probit link)

Assume that the latent random variables $r_{i,j} \sim \mathcal{N}(\boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1)$ and $s_{i,j,k} \sim \mathcal{N}(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i, 1),\ \forall(i,j)$ such that $z_{i,j} = 1$, if $r_{i,j} > 0$; $z_{i,j} = 0$, if $r_{i,j} \leq 0$; $y_{i,j,k} = 1$, if $s_{i,j,k} > 0,\ z_{i,j} = 1$ and $y_{i,j,k} = 0$ if $s_{i,j,k} \leq 0$. The simulated $s_{i,j,k}$ values are stored in a $\sum_{j=1}^{J} K_j \times n_s$ matrix $\boldsymbol{S}$ where $\boldsymbol{S} = [\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_s}]$ and $\boldsymbol{s}_i = [[s_{i,1,1}, \ldots, s_{i,1,K_1}], [s_{i,2,1}, \ldots, s_{i,2,K_2}] \ldots, [s_{i,J,1}, \ldots, s_{i,J,K_J}]]^T$ while the simulated $r_{i,j}$ values are stored in a $n_s \times J$ matrix $\boldsymbol{R}$. Further define $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{s}}_i$ as the elements of $\boldsymbol{V}$ and $\boldsymbol{s}_i$ associated with $z_{i,j} = 1,\ \forall j = 1, \ldots, J$.

$$z_{i,j} = \begin{cases} 1 & \text{if } r_{i,j} > 0, \\ 0 & \text{if } r_{i,j} \leq 0 \end{cases}$$

and

$$y_{i,j,k} = \begin{cases} 1 & \text{if } s_{i,j,k} > 0,\ z_{i,j} = 1 \\ 0 & \text{if } s_{i,j,k} \leq 0,\ z_{i,j} = 0 \text{ or } z_{i,j} = 1. \end{cases}$$

In this case, the joint posterior distribution is proportional to

$$\begin{aligned} &\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}, \nabla_{\boldsymbol{\alpha}}, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}, \nabla_{\boldsymbol{\beta}}, \boldsymbol{z}, \boldsymbol{r}, \boldsymbol{s}|\boldsymbol{y}) \\ &\propto \pi(\boldsymbol{\mu}_{\boldsymbol{\alpha}})\pi(\sigma^2_{\boldsymbol{\alpha}}|\nabla_{\boldsymbol{\alpha}})\pi(\nabla_{\boldsymbol{\alpha}})\pi(\boldsymbol{\mu}_{\boldsymbol{\beta}})\pi(\sigma^2_{\boldsymbol{\beta}}|\nabla_{\boldsymbol{\beta}})\pi(\nabla_{\boldsymbol{\beta}}) \prod_{i=1}^{n_s} \left(\pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}})\pi(\boldsymbol{\beta}_i|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}})\right) \times \\ &\prod_{i=1}^{n_s}\prod_{j=1}^{J} 1^{z_{i,j}}_{\{r_{i,j}>0\}} 1^{1-z_{i,j}}_{\{r_{i,j}\leq 0\}} \phi(r_{i,j}, \boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1) \prod_{k=1}^{K_j} 1^{y_{i,j,k}}_{\{s_{i,j,k}>0\}} 1^{1-y_{i,j,k}}_{\{s_{i,j,k}\leq 0\}} \phi(s_{i,j,k}, \boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i, 1) \end{aligned}$$

where $\phi(., \mu, 1)$ is the probability density function of a Gaussian distribution with mean parameter μ and variance parameter 1 while the $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ matrices contains all of the $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ regression effects.

The full conditional distributions required to undertake Gibbs sampling are all of a familiar form and are derived below. In what follows we use the following notation to denote a conditional posterior distribution; $\pi(x|.)$ where a '.' represents all data and variables that one conditions on.

$\underline{\pi(\sigma^2_{\boldsymbol{\alpha}}|.)}$

$$\begin{aligned}
\pi(\sigma^2_{\boldsymbol{\alpha}}|.) &\propto \pi(\sigma^2_{\boldsymbol{\alpha}}|\nabla_{\boldsymbol{\alpha}}) \prod_{i=1}^{n_s} \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}), \\
&\propto \left(\sigma^2_{\boldsymbol{\alpha}}\right)^{-(1/2+1)} e^{-\frac{1}{\nabla_{\boldsymbol{\alpha}}\sigma^2_{\boldsymbol{\alpha}}}} \prod_{i=1}^{n_s} \left(\sigma^2_{\boldsymbol{\alpha}}\right)^{-n_d/2} \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\alpha}}} \left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T \left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right) \text{ such that} \\
\sigma^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1 + n_s n_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}} + \frac{1}{2}\sum_i \left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T \left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right).
\end{aligned}$$

$\underline{\pi(\nabla_{\boldsymbol{\alpha}}|.)}$

$$\begin{aligned}
\pi(\nabla_{\boldsymbol{\alpha}}|.) &\propto \pi(\sigma^2_{\boldsymbol{\alpha}}|\nabla_{\boldsymbol{\alpha}})\pi(\nabla_{\boldsymbol{\alpha}}), \\
&\propto \nabla_{\boldsymbol{\alpha}}^{-1/2} e^{-\frac{1}{\nabla_{\boldsymbol{\alpha}}\sigma^2_{\boldsymbol{\alpha}}}} \nabla_{\boldsymbol{\alpha}}^{-(1/2+1)} e^{-\frac{1}{A^2\nabla_{\boldsymbol{\alpha}}}} \text{ such that} \\
\nabla_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \frac{1}{\sigma^2_{\boldsymbol{\alpha}}} + \frac{1}{A^2}\right).
\end{aligned}$$

$\underline{\pi(\sigma^2_{\boldsymbol{\beta}}|.)}$

$$\begin{aligned}
\pi(\sigma^2_{\boldsymbol{\beta}}|.) &\propto \pi(\sigma^2_{\boldsymbol{\beta}}|\nabla_{\boldsymbol{\beta}}) \prod_{i=1}^{n_s} \pi(\boldsymbol{\beta}_i|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}), \\
&\propto \left(\sigma^2_{\boldsymbol{\beta}}\right)^{-(1/2+1)} e^{-\frac{1}{\nabla_{\boldsymbol{\beta}}\sigma^2_{\boldsymbol{\beta}}}} \prod_{i=1}^{n_s} \left(\sigma^2_{\boldsymbol{\beta}}\right)^{-n_o/2} \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\beta}}} \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right), \\
&\propto \left(\sigma^2_{\boldsymbol{\beta}}\right)^{-(1/2+n_s n_o/2+1)} \exp\left(-\frac{1}{\nabla_{\boldsymbol{\beta}}\sigma^2_{\boldsymbol{\beta}}} - \frac{1}{2\sigma^2_{\boldsymbol{\beta}}} \sum_{i=1}^{n_s} \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right) \text{ such that} \\
\sigma^2_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(\frac{1}{2}(1 + n_s n_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}} + \frac{1}{2}\sum_{i=1}^{n_s} \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T \left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right).
\end{aligned}$$

$\underline{\pi(\nabla_{\boldsymbol{\beta}}|.)}$

$$\begin{aligned}\pi(\nabla_{\boldsymbol{\beta}}|.) &\propto \pi(\sigma^2_{\boldsymbol{\beta}}|\nabla_{\boldsymbol{\beta}})\pi(\nabla_{\boldsymbol{\beta}}),\\ &\propto \nabla_{\boldsymbol{\beta}}^{-1/2} e^{-\frac{1}{\nabla_{\boldsymbol{\beta}}\sigma^2_{\boldsymbol{\beta}}}} \nabla_{\boldsymbol{\beta}}^{-(1/2+1)} e^{-\frac{1}{B^2\nabla_{\boldsymbol{\beta}}}} \text{ such that}\\ \nabla_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(1,\ \frac{1}{\sigma^2_{\boldsymbol{\beta}}} + \frac{1}{B^2}\right).\end{aligned}$$

$\underline{\pi(\boldsymbol{\mu}_{\boldsymbol{\alpha}}|.) \text{ and } \pi(\boldsymbol{\mu}_{\boldsymbol{\beta}}|.)}$

$$\begin{aligned}\pi(\boldsymbol{\mu}_{\boldsymbol{\alpha}}|.) &\propto \pi(\boldsymbol{\mu}_{\boldsymbol{\alpha}}) \prod_{i=1}^{n_s} \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}),\\ &\propto \exp\left(-\frac{1}{2a^2}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right) \prod_{i=1}^{n_s} \exp\left(-\frac{1}{2\sigma^2_{\boldsymbol{\alpha}}}\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right),\\ &\propto \exp\left(-\frac{1}{2a^2}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \frac{1}{2\sigma^2_{\boldsymbol{\alpha}}}\sum_{i=1}^{n_s}\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right). \qquad (5.4)\end{aligned}$$

We now set about to complete the square in terms of $\boldsymbol{\alpha}_i$. Note that

$$\sum_i \left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right) = n_s\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} - 2\sum_i \boldsymbol{\alpha}_i^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \text{terms independent of } \boldsymbol{\mu}_{\boldsymbol{\alpha}}$$

(denoted as '...' below). The exponent of the exponential in equation (5.4) simplifies as follows

$$\begin{aligned}&-\frac{1}{2}\left(\frac{1}{a^2}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}\left(n_s\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} - 2\sum_i \boldsymbol{\alpha}_i^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \ldots\right)\right)\\ =&-\frac{1}{2}\left(\frac{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}{\sigma^2_{\boldsymbol{\alpha}} a^2}\right)\left(\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} - \frac{2a^2}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\sum_i \boldsymbol{\alpha}_i^T\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \ldots\right)\end{aligned}$$

such that

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. \sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\sum_{i=1}^{n_s}\boldsymbol{\alpha}_i,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\boldsymbol{I}_{n_d}\right).$$

Similarly,

$$\boldsymbol{\mu}_{\boldsymbol{\beta}}|. \sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\sum_{i=1}^{n_s}\boldsymbol{\beta}_i,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\boldsymbol{I}_{n_o}\right).$$

$\underline{\pi(r_{i,j}|.)}$

$$\pi(r_{i,j}|.) \propto 1_{\{r_{i,j}>0\}}^{z_{i,j}} 1_{\{r_{i,j}\leq 0\}}^{1-z_{i,j}} \phi(r_{i,j}, \boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1), \forall (i,j),$$

such that

$$\pi(r_{i,j}|.) \propto \begin{cases} 1_{\{r_{i,j}>0\}} \phi(r_{i,j}, \boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 1, \\ 1_{\{r_{i,j}\leq 0\}} \phi(r_{i,j}, \boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 0. \end{cases}$$

The conditional posterior distribution of $r_{i,j}|.$ is thus a truncated Gaussian distribution,

$$r_{i,j}|. \sim \begin{cases} \mathcal{N}_+(\boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 1, \\ \mathcal{N}_-(\boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 0. \end{cases}$$

The simulated $r_{i,j}$ values are stored in a $n_s \times J$ matrix $\boldsymbol{R}$ where the i^{th} row of $\boldsymbol{R}$ is denoted as $\boldsymbol{r}_i^T$.

$\underline{\pi(\boldsymbol{\beta}_i|.)}$

$$\begin{aligned} \pi(\boldsymbol{\beta}_i|.) &\propto \pi(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \sigma_\beta^2) \prod_{j=1}^{J} \phi(r_{i,j}, \boldsymbol{x}_j^T \boldsymbol{\beta}_i, 1), \\ &\propto \exp\left(-\frac{1}{2\sigma_\beta^2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)^T(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\right) \prod_{j=1}^{J} \exp\left(-\frac{1}{2}(r_{i,j} - \boldsymbol{x}_j^T \boldsymbol{\beta}_i)^2\right), \\ &\propto \exp\left(-\frac{1}{2\sigma_\beta^2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)^T(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\right) \exp\left(-\frac{1}{2}(\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{\beta}_i)^T(\boldsymbol{r}_i - \boldsymbol{X}\boldsymbol{\beta}_i)\right) \end{aligned}$$

and after completing the square in terms of $\boldsymbol{\beta}_i$ we obtain

$$\boldsymbol{\beta}_i|. \sim \mathcal{N}\left(\left(\tau_\beta \boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\tau_\beta \boldsymbol{\mu}_\beta + \boldsymbol{X}^T\boldsymbol{r}_i\right), \left(\tau_\beta \boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right)$$

where $\tau_\beta = \dfrac{1}{\boldsymbol{\sigma}_\beta^2}$.

$\underline{\pi(\boldsymbol{s}_{i,j,k}|.)}$

The conditional posterior distribution of $s_{i,j,k|.}$ only has to be updated when $z_{i,j} = 1$. The case $\{z_{i,j} = 0,\ y_{i,j,k} = 1\}$ is not possible since we assume no false identification of species while the case $\{z_{i,j} = 0,\ y_{i,j,k} = 0,\ \forall k\}$ leads to $\pi(\boldsymbol{\alpha}|.) \propto \prod_{i=1}^{n_s} \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma_{\boldsymbol{\alpha}}^2)$. i.e. the

conditional posterior distribution of $\boldsymbol{\alpha}|.$ is only dependent on the prior distributions of $\boldsymbol{\alpha}_i$.

$$\pi(s_{i,j,k}|.) \propto 1_{\{s_{i,j,k}>0\}}^{y_{i,j,k}} 1_{\{s_{i,j,k}\leq 0\}}^{1-y_{i,j,k}} \phi(s_{i,j,k}, \boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i, 1), \forall (j,k), \tag{5.5}$$

such that the conditional posterior distribution of $s_{i,j,k}|.$ is a truncated Gaussian distribution,

$$s_{i,j,k}|. \sim \begin{cases} \mathcal{N}_+(\boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j,k} = 0, \\ \mathcal{N}_-(\boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j,k} = 1. \end{cases}$$

$\underline{\pi(\boldsymbol{\alpha}_i|.)}$

The posterior distribution of $\boldsymbol{\alpha}_i|.$ only has to be obtained using information when $z_{i,j} = 1$.

Define $\tilde{\boldsymbol{V}}$ and $\tilde{\boldsymbol{s}}_i$ as the elements of $\boldsymbol{V}$ and $\boldsymbol{s}_i$ associated with $z_{i,j} = 1$. We now have

$$\begin{aligned} \pi(\boldsymbol{\alpha}_i|.) &\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma_{\boldsymbol{\alpha}}^2) \prod_{\{j:z_{i,j}=1\}} \prod_k 1_{\{s_{i,j,k}>0\}}^{y_{i,j,k}} 1_{\{s_{i,j,k}\leq 0\}}^{1-y_{i,j,k}} \exp\left(-0.5(s_{i,j,k} - \boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_{\boldsymbol{\alpha}}^2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right) \exp\left(-\frac{1}{2}(\tilde{\boldsymbol{s}}_i - \tilde{\boldsymbol{V}}\boldsymbol{\alpha}_i)^T(\tilde{\boldsymbol{s}}_i - \tilde{\boldsymbol{V}}\boldsymbol{\alpha}_i)\right). \end{aligned}$$

After completing the square in terms of $\boldsymbol{\alpha}_i$ we obtain

$$\boldsymbol{\alpha}_i|. \sim \mathcal{N}\left(\left(\tau_{\boldsymbol{\alpha}} \boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T \tilde{\boldsymbol{V}}\right)^{-1} \left(\tau_{\boldsymbol{\alpha}} \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \tilde{\boldsymbol{V}}^T \tilde{\boldsymbol{s}}_i\right),\ \left(\tau_{\boldsymbol{\alpha}} \boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T \tilde{\boldsymbol{V}}\right)^{-1}\right)$$

where $\tau_{\boldsymbol{\alpha}} = \dfrac{1}{\boldsymbol{\sigma}_{\boldsymbol{\alpha}}^2}$.

$\underline{\pi(z_{i,j}|.)}$

If species i is observed at any of the sites we have $z_{i,j} = 1$ however, if $\boldsymbol{y}_i = \boldsymbol{0}$

$$z_{i,j}|\boldsymbol{y}_i = \boldsymbol{0} \sim \text{Bernoulli}\left(\frac{\Phi(\boldsymbol{x}_j^T \boldsymbol{\beta}_i) \prod_{k=1}^{K_j}(1 - \Phi(\boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i))}{1 - \Phi(\boldsymbol{x}_j^T \boldsymbol{\beta}_i) + \Phi(\boldsymbol{x}_j^T \boldsymbol{\beta}_i) \prod_{k=1}^{K_j}(1 - \Phi(\boldsymbol{v}_{j,k}^T \boldsymbol{\alpha}_i))}\right).$$

The above conditional posterior distributions can be summarised as follows

$$
\begin{aligned}
\sigma^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}}+\frac{1}{2}\sum_{i=1}^{n_s}\left(\boldsymbol{\alpha}_i-\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}_i-\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right),\\
\nabla_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}}+\frac{1}{A^2}\right),\\
\sigma^2_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}}+\frac{1}{2}\sum_{i=1}^{n_s}\left(\boldsymbol{\beta}_i-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{\beta}_i-\boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right),\\
\nabla_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\beta}}+\frac{1}{B^2}\right),\\
\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. &\sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}}+n_s a^2}\sum_{i=1}^{n_s}\boldsymbol{\alpha}_i,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}}+n_s a^2}\boldsymbol{I}_{n_d}\right),\\
\boldsymbol{\mu}_{\boldsymbol{\beta}}|. &\sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}}+n_s b^2}\sum_{i=1}^{n_s}\boldsymbol{\beta}_i,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}}+n_s b^2}\boldsymbol{I}_{n_o}\right),\\
r_{i,j}|. &\sim \begin{cases}\mathcal{N}_+(\boldsymbol{x}_j^T\boldsymbol{\beta}_i,1) & \text{if } z_{i,j}=1,\\ \mathcal{N}_-(\boldsymbol{x}_j^T\boldsymbol{\beta}_i,1) & \text{if } z_{i,j}=0,\end{cases}\\
\boldsymbol{\beta}_i|. &\sim \mathcal{N}\left(\left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o}+\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\tau_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}}+\boldsymbol{X}^T\boldsymbol{r}_i\right),\ \left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o}+\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right),\\
s_{i,j,k}|. &\sim \begin{cases}\mathcal{N}_+(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i,1) & \text{if } z_{i,j}=1,\ y_{i,j,k}=0,\\ \mathcal{N}_-(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i,1) & \text{if } z_{i,j}=1,\ y_{i,j,k}=1,\end{cases}\\
\boldsymbol{\alpha}_i|. &\sim \mathcal{N}\left(\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}+\tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}+\tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{s}}_i\right),\ \left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}+\tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\right),\\
z_{i,j}|. &\sim \begin{cases}\text{Bernoulli}(1) & \text{if } \boldsymbol{y}_i\neq\boldsymbol{0},\\ \text{Bernoulli}\left(\dfrac{\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)\prod_{k=1}^{K_j}(1-\Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i))}{1-\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)+\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)\prod_{k=1}^{K_j}(1-\Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i))}\right) & \text{if } \boldsymbol{y}_i=\boldsymbol{0}\end{cases}
\end{aligned}
$$

where $\tau_{\boldsymbol{\alpha}}=\dfrac{1}{\sigma^2_{\boldsymbol{\alpha}}}$ and $\tau_{\boldsymbol{\beta}}=\dfrac{1}{\sigma^2_{\boldsymbol{\beta}}}$ respectively.

5.A.2 No survey-specific covariates are available (Binomial detection process (probit link)).

Assuming that no survey-specific covariate information is available, detection-nondetection data across surveys at a site could be grouped together to obtain the observed number of times a species was observed at each site (i.e. $y_{i,j}=\sum_{k=1}^{K_j}y_{i,j,k}$). The observed number of species detections at each site could thus be modelled using a binomial distribution

(as suggested by Dorazio and Royle (2005); refer to Section 5.3) if we assume that the conditional detection probabilities are constant across surveys (i.e. $p_{i,j,k} = p_{i,j},\ \forall k$). Using this formulation obtaining a simple known conditional posterior distribution for the detection regression effects is cumbersome. A Gibbs sampling algorithm could be developed for this model; however, the algorithm leads to a dramatic increase in the number of latent variables required when using the Albert and Chib (1993) data-augmentation method. Alternatively, we could also resort to using either importance sampling (Robert and Casella, 1999) or a Metropolis Hastings step to obtain posterior samples for the detection regression effects.

If we however do not perform the grouping across surveys and assume that the conditional detection probabilities are constant across surveys, the detection process can then be modelled as follows,

$$y_{i,j,k}|z_{i,j}, p_{i,j,k} \sim \text{Bernoulli}(z_{i,j}p_{i,j,k}) \text{ with } p_{i,j,k} = p_{i,j} = \Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i),\ \forall k = 1, \ldots, K_j.$$

The $\boldsymbol{V}$ matrix now changes to contain site-specific covariates that could be used to model the detection process and is defined as,

$$\boldsymbol{V} = \begin{bmatrix} \text{site-specific covariates at site 1} \\ \vdots \\ \text{site-specific covariates at site J} \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1^T \\ \vdots \\ \boldsymbol{v}_J^T \end{bmatrix} \tag{5.6}$$

while $\boldsymbol{s}_i^T$ is defined as

$$\boldsymbol{s}_i^T = \begin{bmatrix} \boldsymbol{s}_{i,1}, \boldsymbol{s}_{i,2}, \ldots, \boldsymbol{s}_{i,J} \end{bmatrix}. \tag{5.7}$$

The conditional posterior distributions of $s_{i,j}|.$, $\boldsymbol{\alpha}_i|.$ and $z_{i,j}|\boldsymbol{y}_i, .$ in this case are

$$s_{i,j}|. \sim \begin{cases} N_+(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j} = 0, \\ N_-(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j} = 1, \end{cases}$$

$$\boldsymbol{\alpha}_i|. \sim N\left(\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{s}}_i\right)^T,\ \left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\right),$$

$$z_{i,j}|. \sim \begin{cases} \text{Bernoulli}(1) & \text{if } \boldsymbol{y}_i \neq \mathbf{0}, \\ \text{Bernoulli}\left(\dfrac{\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}{1-\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)+\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}\right) & \text{if } \boldsymbol{y}_i = \mathbf{0}, \end{cases}$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$.

The full set of conditional posterior distributions are shown below. The derivation of the conditional distributions in this case is very similar to what was done in Appendix 5.A.1.

$$\begin{aligned}
\sigma^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right), \\
\nabla_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}} + \frac{1}{A^2}\right), \\
\sigma^2_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right), \\
\nabla_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\beta}} + \frac{1}{B^2}\right), \\
\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. &\sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\sum_{i=1}^{n_s}\boldsymbol{\alpha}_i,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\boldsymbol{I}_{n_d}\right), \\
\boldsymbol{\mu}_{\boldsymbol{\beta}}|. &\sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\sum_{i=1}^{n_s}\boldsymbol{\beta}_i,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\boldsymbol{I}_{n_o}\right), \\
r_{i,j}|. &\sim \begin{cases} \mathcal{N}_+(\boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 1, \\ \mathcal{N}_-(\boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1) & \text{if } z_{i,j} = 0, \end{cases} \\
\boldsymbol{\beta}_i|. &\sim \mathcal{N}\left(\left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\tau_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}} + \boldsymbol{X}^T\boldsymbol{r}_i\right),\ \left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right), \\
s_{i,j,k}|. &\sim \begin{cases} \mathcal{N}_+(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j,k} = 0, \\ \mathcal{N}_-(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i, 1) & \text{if } z_{i,j} = 1,\ y_{i,j,k} = 1, \end{cases} \\
\boldsymbol{\alpha}_i|. &\sim \mathcal{N}\left(\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}^T + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{s}}_i\right)^T,\ \left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\right), \\
z_{i,j}|. &\sim \begin{cases} \text{Bernoulli}(1) & \text{if } \boldsymbol{y}_i \neq \mathbf{0}, \\ \text{Bernoulli}\left(\dfrac{\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}{1-\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)+\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}\right) & \text{if } \boldsymbol{y}_i = \mathbf{0}, \end{cases}
\end{aligned}$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$ and $\tau_{\boldsymbol{\beta}} = \frac{1}{\sigma^2_{\boldsymbol{\beta}}}$ respectively.

5.B Conditional distributions required to undertake a Gibbs sampler for a MSO model when species richness is unknown.

5.B.1 Binomial detection process (probit link)

The derivation of the conditional distributions in this case is very similar to what was done in Appendix 5.A.2. We however require the derivation of the conditional distribution of $w_i|.$ and $\Omega|.$ as well since they were not included in the previous analysis.

$\underline{\pi(w_i|.)}$

Since n_s species are observed, $w_i = 1$ for $i = 1, \ldots, n_s$. We also know that $w_i = 1$ if $z_{i,j} = 1$ for any $j = 1, \ldots, J$ and thus require the conditional distribution of $w_i|.$ for $i > n_s$ where $\boldsymbol{y}_i = \mathbf{0}$ and $z_{i,j} = 0 \; \forall j$. In this case

$$\begin{aligned}\pi(w_i|., i > n_s) &\propto \pi(w_i|\Omega) \prod_{\{j:z_{i,j}=0\}} [z_{i,j}|\psi_{i,j}, w_i][y_{i,j}|p_{i,j}, z_{i,j}, w_i] \\ &\propto \Omega^{w_i}(1-\Omega)^{1-w_i} \prod_{\{j:z_{i,j}=0\}} (1 - \psi_{i,j}w_i)\end{aligned}$$

such that

$$w_i|., i > n_s \sim \text{Bernoulli}\left(\frac{\Omega \prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}{1-\Omega+\Omega\prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}\right) \text{ if } z_{i,j} = 0 \; \forall j, \; i > n_s. \quad (5.8)$$

$\underline{\pi(\Omega|.)}$

Assume that the prior distribution of Ω is a Beta distribution with parameters α_Ω and β_Ω. In this case the posterior distribution for $\Omega|.$ is a Beta distribution with parameters $\alpha_\Omega + \sum_i w_i$ and $S + \beta_\Omega + \sum_i w_i$.

We display the full set of conditional posterior distributions below.

$$w_i|. \sim \begin{cases} \text{Bernoulli}(1) & \text{if } i \leq n_s, \\ \text{Bernoulli}(1) & \text{if } z_{i,j} = 1 \text{ for any } j,\ i > n_s, \\ \text{Bernoulli}\left(\dfrac{\Omega \prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}{1-\Omega+\Omega\prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}\right) & \text{if } \boldsymbol{y}_i = \boldsymbol{0},\ z_{i,j}=0,\ \forall j,\ i > n_s, \end{cases}$$

$$\Omega|. \sim \text{Beta}(\alpha_\Omega + \sum_i w_i,\ S + \beta_\Omega + \sum_i w_i),$$

$$\sigma^2_{\boldsymbol{\alpha}}|. \sim \text{IG}\left(\frac{1}{2}(1+Sn_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right),$$

$$\nabla_{\boldsymbol{\alpha}}|. \sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}} + \frac{1}{A^2}\right),$$

$$\sigma^2_{\boldsymbol{\beta}}|. \sim \text{IG}\left(\frac{1}{2}(1+Sn_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right),$$

$$\nabla_{\boldsymbol{\beta}}|. \sim \text{IG}\left(1,\ \tau_{\boldsymbol{\beta}} + \frac{1}{B^2}\right),$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. \sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}} + Sa^2}\sum_i \boldsymbol{\alpha}_i^T,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}} + Sa^2}\boldsymbol{I}_{n_d}\right),$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}}|. \sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}} + Sb^2}\sum_i \boldsymbol{\beta}_i^T,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}} + Sb^2}\boldsymbol{I}_{n_o}\right),$$

$$r_{i,j}|. \sim \begin{cases} \mathcal{N}_+(\boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1) & \text{if } w_i = 1,\ z_{i,j} = 1, \\ \mathcal{N}_-(\boldsymbol{x}_j^T\boldsymbol{\beta}_i, 1) & \text{if } w_i = 1,\ z_{i,j} = 0, \end{cases}$$

$$\boldsymbol{\beta}_i|. \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o}) & \text{if } w_i = 0, \\ \mathcal{N}\left(\left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\left(\tau_{\boldsymbol{\beta}}\boldsymbol{\mu}_{\boldsymbol{\beta}} + \boldsymbol{X}^T\boldsymbol{r}_i\right),\ \left(\tau_{\boldsymbol{\beta}}\boldsymbol{I}_{n_o} + \boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\right) & \text{if } w_i = 1, \end{cases}$$

$$\pi(\boldsymbol{\alpha}_i|.) \propto \begin{cases} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}) & \text{if } w_i = 0, \\ \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}})\prod_{\{j:z_{i,j}=1\}}^{J}\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i)^{y_{i,j}}\left(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i)\right)^{K_j - y_{i,j}} & \text{if } w_i = 1, \end{cases}$$

$$z_{i,j}|. \sim \begin{cases} \text{Bernoulli}(0) & \text{if } w_i = 0, \\ \text{Bernoulli}\left(\dfrac{\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}{1-\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)+\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)(1-\Phi(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i))^{K_j}}\right) & \text{if } w_i = 1,\ \boldsymbol{y}_i = \boldsymbol{0}, \\ \text{Bernoulli}(1) & \text{if } w_i = 1,\ \boldsymbol{y}_i \neq \boldsymbol{0} \end{cases}$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$ and $\tau_{\boldsymbol{\beta}} = \frac{1}{\sigma^2_{\boldsymbol{\beta}}}$ respectively.

Take note that $r_{i,j}|.\forall j$ is not sampled when $w_i = 0$.

5.B.2 Bernoulli detection process (probit link)

If the detection process was defined as $y_{i,j,k}|z_{i,j}, p_{i,j,k} \sim \text{Bernoulli}(w_i z_{i,j} p_{i,j,k})$ and survey specific covariate information is available, then the posterior distribution of the parameters of the model is

$$\begin{aligned}
&\pi(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}, \nabla_{\boldsymbol{\alpha}}, \boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}}, \nabla_{\boldsymbol{\beta}}, \boldsymbol{z}, \boldsymbol{w}, \Omega|\boldsymbol{y}) \\
&\propto \pi(\boldsymbol{\mu}_{\boldsymbol{\alpha}})\pi(\sigma^2_{\boldsymbol{\alpha}}|\nabla_{\boldsymbol{\alpha}})\pi(\nabla_{\boldsymbol{\alpha}})\pi(\boldsymbol{\mu}_{\boldsymbol{\beta}})\pi(\sigma^2_{\boldsymbol{\beta}}|\nabla_{\boldsymbol{\beta}})\pi(\nabla_{\boldsymbol{\beta}})\pi(\Omega) \prod_{i=1}^{S} \left(\pi(w_i|\Omega)\pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}})\pi(\boldsymbol{\beta}_i|\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma^2_{\boldsymbol{\beta}})\right) \times \\
&\quad \prod_{i=1}^{S}\prod_{j=1}^{J} \left(\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)^{z_{i,j}} \left(1 - \Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)\right)^{1-z_{i,j}} \prod_{k=1}^{K_j} (z_{i,j}\Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i))^{y_{i,j,k}} \left(1 - z_{i,j}\Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i)\right)^{1-y_{i,j,k}} \right)^{w_i}.
\end{aligned}$$

Most of the conditional posterior distributions as defined in Appendix 5.A.1 remains unchanged except for the distributions of $\boldsymbol{\alpha}_i|.$ and $z_{i,j}|.$. Since the observation process is assumed to be a Bernoulli random variable, data augmentation can be used by introducing the $s_{i,j,k}$ variables as done previously. The conditional posterior distributions of $\boldsymbol{\alpha}_i|.$ and $z_{i,j}|.$ and $s_{i,j,k}|.$ are now,

$$s_{i,j,k}|. \sim \begin{cases} \mathcal{N}_+(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i, 1) & \text{if } w_i = 1,\ z_{i,j} = 1,\ y_{i,j,k} = 0, \\ \mathcal{N}_-(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i, 1) & \text{if } w_i = 1, z_{i,j} = 1,\ y_{i,j,k} = 1, \end{cases}$$

$$\boldsymbol{\alpha}_i|. \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d}) & \text{if } w_i = 0, \\ \mathcal{N}\left(\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\left(\tau_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \tilde{\boldsymbol{V}}\tilde{\boldsymbol{s}}_i{}^T\right),\ \left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\tilde{\boldsymbol{V}}\right)^{-1}\right) & \text{if } w_i = 1, \end{cases}$$

$$z_{i,j}|. \sim \begin{cases} \text{Bernoulli}(0) & \text{if } w_i = 0, \\ \text{Bernoulli}\left(\dfrac{\Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)\prod_k(1 - \Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i))}{1 - \Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i) + \Phi(\boldsymbol{x}_j^T\boldsymbol{\beta}_i)\prod_k(1 - \Phi(\boldsymbol{v}_{j,k}^T\boldsymbol{\alpha}_i))}\right) & \text{if } w_i = 1,\ \boldsymbol{y}_i = \boldsymbol{0}, \\ \text{Bernoulli}(1) & \text{if } w_i = 1,\ \boldsymbol{y}_i \neq \boldsymbol{0} \end{cases}$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$. Note $\tilde{\boldsymbol{s}}_i$ and $\tilde{\boldsymbol{V}}$ are defined as per Section 5.A.1

5.C Conditional distributions required to undertake a Gibbs sampler for a MSO model when species richness is unknown (Binomial detection process (logistic link)).

Here we derive the conditional posterior distributions required to sample from $\boldsymbol{\alpha}_i|.$ and $\boldsymbol{\beta}_i|.$ respectively.

$\underline{\boldsymbol{\alpha}_i|.}$

When $w_i = 0$, species i does not occupy the region under investigation and $\underline{\boldsymbol{\alpha}_i|.}$ is not updated.

When $w_i = 1$ we have,

$$\begin{aligned}
\pi(\boldsymbol{\alpha}_i|w_i = 1, .) &\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}) \prod_j [y_{i,j}|z_{i,j}, p_{i,j}] \\
&\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}) \prod_j (z_{i,j}p_{i,j})^{y_{i,j}} (1 - z_{i,j}p_{i,j})^{K_j - y_{i,j}} \\
&\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}) \prod_{\{j:z_{i,j}=1\}} p_{i,j}^{y_{i,j}} (1 - p_{i,j})^{K_j - y_{i,j}} \\
&\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}) \prod_{\{j:z_{i,j}=1\}} \left(\frac{p_{i,j}}{1 - p_{i,j}}\right)^{y_{i,j}} (1 - p_{i,j})^{K_j} \\
&\propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma^2_{\boldsymbol{\alpha}}) \prod_{\{j:z_{i,j}=1\}} \frac{\left(e^{\boldsymbol{v}_j^T \boldsymbol{\alpha}_i}\right)^{y_{i,j}}}{\left(1 + e^{\boldsymbol{v}_j^T \boldsymbol{\alpha}_i}\right)^{K_j}}.
\end{aligned}$$

Polson et al. (2013) show that a binomial likelihood can be related to Pólya-Gamma such that,

$$\frac{(e^{\psi})^a}{(1 + e^{\psi})^b} = 2^{-b} e^{\kappa\psi} \int_0^{\infty} e^{-\omega\psi^2/2} p(\omega) d\omega, \tag{5.9}$$

where $a \in \mathbb{R}$, $\omega \sim PG(b, 0)$, $b > 0$, $\kappa = a - b/2$.

From equation (5.9) it follows that

$$\pi(\boldsymbol{\alpha}_i|w_i=1,.) \propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}},\sigma^2_{\boldsymbol{\alpha}}) \prod_{\{j:z_{i,j}=1\}} e^{\theta^{(\boldsymbol{\alpha})}_{i,j}\boldsymbol{v}_j^T\boldsymbol{\alpha}_i} \int_0^\infty e^{-\omega^{(\boldsymbol{\alpha})}_{i,j}(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i)^2/2} p(\omega^{(\boldsymbol{\alpha})}_{i,j}) d\omega^{(\boldsymbol{\alpha})}_{i,j}$$

where $\theta^{(\boldsymbol{\alpha})}_{i,j} = y_{i,j} - K_j/2$.

If one now conditions on the values of $\boldsymbol{\omega}^{(\boldsymbol{\alpha})}_i = \left[\omega^{(\boldsymbol{\alpha})}_{i,1},\ldots,\omega^{(\boldsymbol{\alpha})}_{i,J}\right]^T$ then

$$\pi(\boldsymbol{\alpha}_i|w_i=1,.) \propto \pi(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_{\boldsymbol{\alpha}},\sigma^2_{\boldsymbol{\alpha}}) \prod_{\{j:z_{i,j}=1\}} e^{\theta^{(\boldsymbol{\alpha})}_{i,j}\boldsymbol{v}_j^T\boldsymbol{\alpha}_i - \omega^{(\boldsymbol{\alpha})}_{i,j}(\boldsymbol{v}_j^T\boldsymbol{\alpha}_i)^2/2} \tag{5.10}$$

and after completing the square in terms of α_i we obtain,

$$\boldsymbol{\alpha}_i|\boldsymbol{y},\boldsymbol{\omega}^{(\boldsymbol{\alpha})}_i \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i})$$

with

$$\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i} = \left(\tau_{\boldsymbol{\alpha}}\boldsymbol{I}_{n_d} + \tilde{\boldsymbol{V}}^T\boldsymbol{D}_{\boldsymbol{\alpha}}\tilde{\boldsymbol{V}}\right)^{-1}$$
$$\boldsymbol{\mu}_{\boldsymbol{\alpha}_i} = \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i}\left(\tilde{\boldsymbol{V}}^T\boldsymbol{\Theta}_{\boldsymbol{\alpha}} + \tau_{\boldsymbol{\alpha}}\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)$$

and $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$. The matrices $\boldsymbol{D}_{\boldsymbol{\alpha}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\alpha}}$ are defined as, $\boldsymbol{D}_{\boldsymbol{\alpha}} = \text{diagonal}\left[\omega^{(\boldsymbol{\alpha})}_{i,1},\ldots,\omega^{(\boldsymbol{\alpha})}_{i,j^*}\right]$ and $\boldsymbol{\Theta}_{\boldsymbol{\alpha}} = [y_{i,1} - K_1/2,\ldots,y_{i,j^*} - K_{j^*}/2]^T$.

$\underline{\omega^{(\boldsymbol{\alpha})}_{i,j}|w_i=1,.}$

From equation (5.10) it follows that $\omega^{(\boldsymbol{\alpha})}_{i,j}|w_i=1,. \sim \text{PG}(K_j, \boldsymbol{v}_j^T\boldsymbol{\alpha}_i),\ \forall j$.

$\underline{\boldsymbol{\beta}_i|w_i=1,. \text{ and } \omega^{(\boldsymbol{\beta})}_{i,j}|w_i=1,.}$

The posterior distributions of $\boldsymbol{\beta}_i|w_i=1,.$ and $\omega^{(\boldsymbol{\beta})}_{i,j}|w_i=1,.$ can be obtained similar to what was done for $\boldsymbol{\alpha}_i|w_i=1,.$ and $\omega^{(\boldsymbol{\alpha})}_{i,j}|.,w_i=1$.

Below we display the full set of conditional posterior distributions for the MSO model.

$$w_i|. \sim \begin{cases} \text{Bernoulli}(1) & \text{if } i \leq n_s \\ \text{Bernoulli}(1) & \text{if } z_{i,j} = 1 \text{ for any } j,\ i > n_s \\ \text{Bernoulli}\left(\dfrac{\Omega \prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}{1-\Omega+\Omega\prod_{\{j:z_{i,j}=0\}}(1-\psi_{i,j})}\right) & \text{if } z_{i,j} = 0\ \forall j,\ i > n_s \end{cases}$$

$$\Omega|. \sim \text{Beta}(\alpha_\Omega + \sum_i w_i,\ S + \beta_\Omega + \sum_i w_i)$$

$$\sigma^2_{\boldsymbol{\alpha}}|. \sim \text{IG}\left(\frac{1}{2}(1+Sn_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right)$$

$$\nabla_{\boldsymbol{\alpha}}|. \sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}} + \frac{1}{A^2}\right)$$

$$\sigma^2_{\boldsymbol{\beta}}|. \sim \text{IG}\left(\frac{1}{2}(1+Sn_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}} + \frac{1}{2}\sum_{i=1}^{n_s}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})^T(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}})\right)$$

$$\nabla_{\boldsymbol{\beta}}|. \sim \text{IG}\left(1,\ \tau_{\boldsymbol{\beta}} + \frac{1}{B^2}\right)$$

$$\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. \sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}} + Sa^2}\sum_{i=1}^{n_s}\boldsymbol{\alpha}_i^T,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}} + Sa^2}\boldsymbol{I}_{n_d}\right),$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}}|. \sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}} + Sb^2}\sum_{i=1}^{n_s}\boldsymbol{\beta}_i^T,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}} + Sb^2}\boldsymbol{I}_{n_o}\right),$$

$$\boldsymbol{\omega}^{(\boldsymbol{\beta})}_{i,j}|\boldsymbol{\beta}, w_i = 1, . \sim \text{PG}(1, \boldsymbol{x}_j^T\boldsymbol{\beta}_i),\ \forall j = 1, \ldots, J$$

$$\boldsymbol{\beta}_i|\boldsymbol{y}, \boldsymbol{\omega}^{(\boldsymbol{\beta})}_i, w_i = 1, . \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i})$$

$$\boldsymbol{\omega}^{(\boldsymbol{\alpha})}_{i,j}|w_i = 1, . \sim \text{PG}(K_j, \boldsymbol{v}_j^T\boldsymbol{\alpha}_i),\ \forall j.$$

$$\boldsymbol{\alpha}_i|\boldsymbol{y}, \boldsymbol{\omega}^{(\boldsymbol{\alpha})}_i, w_i = 1, . \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i})$$

$$z_{i,j}|. \sim \begin{cases} \text{Bernoulli}(0) & \text{if } w_i = 0 \\ \text{Bernoulli}\left(\dfrac{\psi_{i,j}(1-p_{i,j})^{K_j}}{1-\psi_{i,j}+\psi_{i,j}(1-p_{i,j})^{K_j}}\right) & \text{if } w_i = 1,\ \boldsymbol{y}_i = \boldsymbol{0} \\ \text{Bernoulli}(1) & \text{if } w_i = 1,\ \boldsymbol{y}_i \neq \boldsymbol{0} \end{cases}$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$ and $\tau_{\boldsymbol{\beta}} = \frac{1}{\sigma^2_{\boldsymbol{\beta}}}$ respectively.

5.D Conditional distributions required to undertake a Gibbs sampler for a MSO model when species richness is known.

5.D.1 Binomial detection process for ungrouped data (logistic link)

If the species richness is known, then the conditional posterior distributions can be shown to be as follows:

$$
\begin{aligned}
\sigma^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}} + \frac{1}{2}\sum_{i=1}^{n_s}\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}_i - \boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right),\\
\nabla_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}} + \frac{1}{A^2}\right),\\
\sigma^2_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_s n_o),\ \frac{1}{\nabla_{\boldsymbol{\beta}}} + \frac{1}{2}\sum_{i=1}^{n_s}\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{\beta}_i - \boldsymbol{\mu}_{\boldsymbol{\beta}}\right)\right),\\
\nabla_{\boldsymbol{\beta}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\beta}} + \frac{1}{B^2}\right),\\
\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. &\sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\sum_i \boldsymbol{\alpha}_i^T,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}} + n_s a^2}\boldsymbol{I}_{n_d}\right),\\
\boldsymbol{\mu}_{\boldsymbol{\beta}}|. &\sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\sum_i \boldsymbol{\beta}_i^T,\ \frac{b^2\sigma^2_{\boldsymbol{\beta}}}{\sigma^2_{\boldsymbol{\beta}} + n_s b^2}\boldsymbol{I}_{n_o}\right),\\
\boldsymbol{\omega}^{(\beta)}_{i,j}|. &\sim \text{PG}(1, \boldsymbol{x}_j^T\boldsymbol{\beta}_i),\ \forall j = 1,\ldots,J,\\
\boldsymbol{\beta}_i|. &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i}),\\
\boldsymbol{\omega}^{(\boldsymbol{\alpha})}_{i,j}|. &\sim \text{PG}(K_j, \boldsymbol{v}_j^T\boldsymbol{\alpha}_i),\ \forall j,\\
\boldsymbol{\alpha}_i|. &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i}),\\
z_{i,j}|. &\sim \begin{cases} \text{Bernoulli}\left(\dfrac{\psi_{i,j}(1-p_{i,j})^{K_j}}{1-\psi_{i,j}+\psi_{i,j}(1-p_{i,j})^{K_j}}\right) & \text{if } \boldsymbol{y}_i = \boldsymbol{0},\\ \text{Bernoulli}(1) & \text{if } \boldsymbol{y}_i \neq \boldsymbol{0}\end{cases}
\end{aligned}
$$

where $\tau_{\boldsymbol{\alpha}} = \frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$ and $\tau_{\boldsymbol{\beta}} = \frac{1}{\sigma^2_{\boldsymbol{\beta}}}$ respectively.

5.D.2 Binomial detection process for grouped data (logistic link)

In the above formulations of the MSOM model it is assumed that the detection and occupancy regression effects are modelled as realisations from common Gaussian distributions (as shown in equations 5.11). Below we add one additional level of complexity and assume that the regression effects of different groups (e.g. bird guilds, taxa or species groups) are modelled as realisations from different Gaussian distributions so as to identify whether or not a group effect exists.

We specifically assume that the regression effects for detection and occupancy of species i in group g are modelled as logistic $\left(p_{i,j}^{(g)}\right) = \boldsymbol{v}_j^T \tilde{\boldsymbol{\alpha}}_i^{(g)}$ and logistic $\left(\psi_{i,j}^{(g)}\right) = \boldsymbol{x}_j^T \tilde{\boldsymbol{\beta}}_i^{(g)}$ where

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{\alpha}} &\sim \mathcal{N}(\mathbf{0}, a^2 \boldsymbol{I}_{n_d}), & \boldsymbol{\mu}_{\boldsymbol{\beta}} &\sim \mathcal{N}(\mathbf{0}, b^2 \boldsymbol{I}_{n_o}) \\ \boldsymbol{\alpha}^{(g)} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}}, \sigma_{\boldsymbol{\alpha}}^2 \boldsymbol{I}_{n_d}), & \boldsymbol{\beta}^{(g)} &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_{\boldsymbol{\beta}}^2 \boldsymbol{I}_{n_o}) \\ \tilde{\boldsymbol{\alpha}}_i^{(g)} &\sim \mathcal{N}(\boldsymbol{\alpha}^{(g)}, \sigma_{\tilde{\boldsymbol{\alpha}}}^2 \boldsymbol{I}_{n_d}), & \tilde{\boldsymbol{\beta}}_i^{(g)} &\sim \mathcal{N}(\tilde{\boldsymbol{\beta}}^{(g)}, \sigma_{\tilde{\boldsymbol{\beta}}}^2 \boldsymbol{I}_{n_0}) \end{aligned} \tag{5.11}$$

for $i = 1, \ldots, n_g;\ j = 1, \ldots, J$ and $g = 1, \ldots, G$. The variance terms in the above set of equations are modelled using a half-Cauchy distribution (Gelman, 2006) and can be specified hierarchically as

$$\begin{aligned} \sigma_{\boldsymbol{\alpha}}^2 | \nabla_{\boldsymbol{\alpha}} &\sim IG\left(0.5, \frac{1}{\nabla_{\boldsymbol{\alpha}}}\right), \nabla_{\boldsymbol{\alpha}} \sim IG\left(0.5, \frac{1}{A^2}\right), \\ \sigma_{\tilde{\boldsymbol{\alpha}}}^2 | \nabla_{\tilde{\boldsymbol{\alpha}}} &\sim IG\left(0.5, \frac{1}{\nabla_{\tilde{\boldsymbol{\alpha}}}}\right), \nabla_{\tilde{\boldsymbol{\alpha}}} \sim IG\left(0.5, \frac{1}{\tilde{A}^2}\right), \\ \sigma_{\boldsymbol{\beta}}^2 | \nabla_{\boldsymbol{\beta}} &\sim IG\left(0.5, \frac{1}{\nabla_{\boldsymbol{\beta}}}\right), \nabla_{\boldsymbol{\beta}} \sim IG\left(0.5, \frac{1}{B^2}\right), \\ \sigma_{\tilde{\boldsymbol{\beta}}}^2 | \nabla_{\tilde{\boldsymbol{\beta}}} &\sim IG\left(0.5, \frac{1}{\nabla_{\tilde{\boldsymbol{\beta}}}}\right), \tilde{\nabla}_{\tilde{\boldsymbol{\beta}}} \sim IG\left(0.5, \frac{1}{\tilde{B}^2}\right) \end{aligned}$$

where A, $\tilde{A}$, B and $\tilde{B}$ are known constants.

Denote all detection regression effects at the species level as $\boldsymbol{\alpha}^{(I)}$ while those at the group level are denoted as $\boldsymbol{\alpha}^{(G)}$. Similarly denote the two sets of occupancy regression effects as $\boldsymbol{\beta}^{(I)}$ and $\boldsymbol{\beta}^{(G)}$.

The resulting conditional probability distributions are as follows:

$$
\begin{aligned}
\sigma^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1+Gn_d),\ \frac{1}{\nabla_{\boldsymbol{\alpha}}}+\frac{1}{2}\sum_{g=1}^{G}\left(\boldsymbol{\alpha}^{(g)}-\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)^T\left(\boldsymbol{\alpha}^{(g)}-\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right),\\
\nabla_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \tau_{\boldsymbol{\alpha}}+\frac{1}{A^2}\right),\\
\tilde{\sigma}^2_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(\frac{1}{2}(1+n_sn_d),\ \frac{1}{\tilde{\nabla}_{\boldsymbol{\alpha}}}+\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\left(\boldsymbol{\alpha}_i^{(g)}-\boldsymbol{\alpha}^{(g)}\right)^T\left(\boldsymbol{\alpha}_i^{(g)}-\boldsymbol{\mu}_{\boldsymbol{\alpha}}\right)\right),\\
\tilde{\nabla}_{\boldsymbol{\alpha}}|. &\sim \text{IG}\left(1,\ \tilde{\tau}_{\boldsymbol{\alpha}}+\frac{1}{\tilde{A}^2}\right),\\
\sigma^2_{\beta}|. &\sim \text{IG}\left(\frac{1}{2}(1+Gn_o),\ \frac{1}{\nabla_{\beta}}+\frac{1}{2}\sum_{g=1}^{G}\left(\boldsymbol{\beta}^{(g)}-\boldsymbol{\mu}_{\beta}\right)^T\left(\boldsymbol{\beta}^{(g)}-\boldsymbol{\mu}_{\beta}\right)\right),\\
\nabla_{\beta}|. &\sim \text{IG}\left(\frac{1}{2},\ \tau_{\beta}+\frac{1}{B^2}\right),\\
\boldsymbol{\mu}_{\boldsymbol{\alpha}}|. &\sim \mathcal{N}\left(\frac{a^2}{\sigma^2_{\boldsymbol{\alpha}}+n_sa^2}\sum_i\boldsymbol{\alpha}_i^T,\ \frac{a^2\sigma^2_{\boldsymbol{\alpha}}}{\sigma^2_{\boldsymbol{\alpha}}+n_sa^2}\boldsymbol{I}_{n_d}\right),\\
\boldsymbol{\mu}_{\beta}|. &\sim \mathcal{N}\left(\frac{b^2}{\sigma^2_{\beta}+n_sb^2}\sum_i\boldsymbol{\beta}_i^T,\ \frac{b^2\sigma^2_{\beta}}{\sigma^2_{\beta}+n_sb^2}\boldsymbol{I}_{n_o}\right),\\
\boldsymbol{\omega}^{(\beta)}_{i,j}|. &\sim \text{PG}(1,\boldsymbol{x}_j^T\boldsymbol{\beta}_i),\ \forall j=1,\ldots,J,\\
\boldsymbol{\beta}_i|. &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\beta}_i}),\\
\boldsymbol{\omega}^{(\boldsymbol{\alpha})}_{i,j}|. &\sim \text{PG}(K_j,\boldsymbol{v}_j^T\boldsymbol{\alpha}_i),\ \forall j,\\
\boldsymbol{\alpha}_i|. &\sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\alpha}_i},\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_i}),\\
z_{i,j}|. &\sim \begin{cases}\text{Bernoulli}\left(\dfrac{\psi_{i,j}(1-p_{i,j})^{K_j}}{1-\psi_{i,j}+\psi_{i,j}(1-p_{i,j})^{K_j}}\right) & \text{if } \boldsymbol{y}_i=\boldsymbol{0},\\ \text{Bernoulli}(1) & \text{if } \boldsymbol{y}_i\neq\boldsymbol{0}\end{cases}
\end{aligned}
$$

where $\tau_{\boldsymbol{\alpha}}=\frac{1}{\sigma^2_{\boldsymbol{\alpha}}}$, $\tilde{\tau}_{\tilde{\boldsymbol{\alpha}}}=\frac{1}{\sigma^2_{\tilde{\boldsymbol{\alpha}}}}$ and $\tau_{\beta}=\frac{1}{\sigma^2_{\beta}}$ respectively.

5.E Species included in the study

Table 19. Species observed at least once at Anysberg with some characteristics

Guild	Latin name	Common name	Number of Observations
Carnivore	*Felis sylvestris*	wildcat	82
	Mellivora capensis	honey badger	16
	Panthera pardus	leopard	21
	Aonyx capensis	African clawless otter	1
	Atilax paludinosus	marsh mongoose	1
	Galerella pulverulenta	Cape grey mongoose	33
	Ictonyx striatus	striped polecat	1
	Canis mesomelas	Black-backed jackal	141
	Caracal caracal	Caracal	12
	Hyaena brunnea	brown hyaena	13
Herbivore	*Oryx gazella*	gemsbok	260
	Taurotragus oryx	common eland	42
	Tragelaphus strepsiceros	greater kudu	52
	Alcelaphus buselaphus caama	red hartebeest	68
	Equus zebra	mountain zebra	13
	Lepus spp	hare	91
	Oreotragus oreotragus	klipspringer	144
	Pelea capreolus	grey rhebok	47
	Procavia capensis	rock hyrax	18
	Pronolagus rupestris	Smith's red rock hare	61
	Raphicerus campestris	steenbok	85
	Sylvicapra grimmia	common duiker	163
	Antidorcas marsupialis	springbok	63
Insectivore	*Orycteropus afer*	aardvark	26
	Proteles cristata	aardwolf	18
	Otocyon megalotis	bat-eared fox	3
Omnivore	*Chlorocebus pygerythrus*	vervet monkey	14
	Genetta genetta	common genet	16
	Hystrix africaeaustralis	Cape porcupine	89
	Papio ursinus	chacma baboon	320
	Potamochoerus larvatus	bushpig	4
Birds	*Eupodotis afra*	Southern black korhaan	31
	Eupodotis vigorsii	Karoo korhaan	15
	Pternistes capensis	Cape spurfowl	14
	Scleroptila africanus	grey-winged francolin	6

5.F Models fitted

Table 20. The covariates that were included in the occupancy and detection processes for the MSO model.

Model	Occupancy covariates	Detection covariates
1	1	1
2	1	trail
3	1	plain + river
4	1	trail + plain + river
5	elevation	1
6	elevation	trail
7	elevation	plain + river
8	elevation	trail + plain + river
9	MSAVI2ind	1
10	MSAVI2ind	trail
11	MSAVI2ind	plain + river
12	MSAVI2ind	trail + plain + river
13	elevation + MSAVI2ind	1
14	elevation + MSAVI2ind	trail
15	elevation + MSAVI2ind	plain + river
16	elevation + MSAVI2ind	trail + plain + river
17	PreyIndex	1
18	PreyIndex	trail
19	PreyIndex	plain + river
20	PreyIndex	trail + plain + river
21	elevation + PreyIndex	1
22	elevation + PreyIndex	trail
23	elevation + PreyIndex	plain + river
24	elevation + PreyIndex	trail + plain + river
25	MSAVI2ind + PreyIndex	1
26	MSAVI2ind + PreyIndex	trail
27	MSAVI2ind + PreyIndex	plain + river
28	MSAVI2ind + PreyIndex	trail + plain + river
29	elevation + MSAVI2ind + PreyIndex	1
30	elevation + MSAVI2ind + PreyIndex	trail
31	elevation + MSAVI2ind + PreyIndex	plain + river
32	elevation + MSAVI2ind + PreyIndex	trail + plain + river

5.G Model selection

Below we briefly highlight different model selection methods that could be used to undertake model selection in a Bayesian context. We specifically focus on WAIC (Watanabe, 2010), the Bayesian p-value (Meng et al., 1994), sampled Bayesian p-value (Zhang, 2014) as well as the reversible jump MCMC algorithm of Barker and Link (2013). DIC (Spiegelhalter et al., 1998) has not been used since several authors suggest that the method is inappropriate for mixture models (Spiegelhalter et al., 2002; Celeux et al., 2006; Plummer, 2008).

WAIC

WAIC is calculated as $\text{WAIC} = -2\text{elppd} + 2\text{p}_{WAIC}$ where *elppd* is the pointwise expected log predictive posterior density and p_{WAIC} is a measure of model complexity (Gelman et al., 2014). As per Broms et al. (2016), we define *elppd* as

$$\text{elppd} \approx \sum_{i=1}^{n_s}\sum_{j=1}^{J} \log\left(\frac{\sum_{s=1}^{S}[\boldsymbol{y}_{i,j}|p_{i,j},\psi_{i,j}]}{S}\right),$$

p_{WAIC} as

$$p_{WAIC} \approx \sum_{i=1}^{n_s}\sum_{j=1}^{J} \text{Var}_{\text{post}}\left(\log[\boldsymbol{y}_{i,j}|p_{i,j},\psi_{i,j}]\right)$$

and the contribution of species i at location j to the integrated likelihood as

$$[\boldsymbol{y}_{i,j}|p_{i,j},\psi_{i,j}] = \psi_{i,j}\binom{K_j}{y_{i,j}}p_{i,j}^{y_{i,j}}(1-p_{i,j})^{K_j-y_{i,j}} + (1-\psi_{i,j})(1-p_{i,j})^{K_j}I_{y_{i,j}=0}. \tag{5.12}$$

In the above formulae, $p_{i,j}$ and $\psi_{i,j}$ are calculated using the post burn-in posterior samples, $\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(S)}$ where $\boldsymbol{\theta}^{(k)}$ is the k^{th} posterior draw from the posterior distribution of the parameters of the MSO model.

p_{WAIC} is unreliably estimated if any of the $\text{Var}_{\text{post}}\left(\log[\boldsymbol{y}_{i,j}|p_{i,j},\psi_{i,j}]\right)$ terms are greater than 0.4 (Vehtari et al., 2017). In a model selection context, the aim is to select a model with the lowest WAIC score or a group of models close to the minimum WAIC score.

Bayesian p-value and sampled Bayesian p-value

The Bayesian p-value relies on the use of a discrepancy function $T(\boldsymbol{y}, \boldsymbol{\theta})$ which is used to compare the data being analysed ($\boldsymbol{y}$) to replicate data ($\boldsymbol{y}^{\text{rep}}$). $\boldsymbol{y}^{\text{rep}}$ is obtained by simulating from $[\boldsymbol{\theta}|\boldsymbol{y}]$ and $\boldsymbol{y}^{\text{rep}} \sim [\boldsymbol{y}|\boldsymbol{\theta}]$ in turn many times. The Bayesian p-value is defined as $\Pr(T(\boldsymbol{y}^{\text{rep}}, \boldsymbol{\theta}) \geq T(\boldsymbol{y}, \boldsymbol{\theta}))$. The measure however is biased towards 0.5 and has low power to detect significant differences in the discrepancy measure. The method should not be used to undertake model selection but rather to highlight whether a model is clearly inappropriate (i.e. when $p < 0.05$ or $p > 0.95$) (Conn et al., 2018).

A sampled Bayesian p-value relies on a single draw from the posterior probability distribution $[\boldsymbol{\theta}|\boldsymbol{y}]$, say $\tilde{\boldsymbol{\theta}}$ and then calculating $\Pr(T(\boldsymbol{y}^{\text{rep}}, \tilde{\boldsymbol{\theta}}) \geq T(\boldsymbol{y}, \tilde{\boldsymbol{\theta}}))$ (Zhang, 2014). Intuitively if $\boldsymbol{y}$ and $\boldsymbol{y}^{\text{rep}}$ are *close* to each other the p-value should be close to 0.5. The variability of the p-value can be obtained using further simulations although this can be prohibitively time-consuming (Hjort et al., 2006). Gosselin (2011) show that asymptotically the distribution of the sampled Bayesian p-value is uniformly distributed over the interval $[0, 1]$ which suggests that misspecified models will not have uniformly distributed sample Bayesian p-values.

Reversible jump MCMC (RJMCMC)

RJMCMC is a Bayesian model selection technique that can be used to obtain the model probabilities associated with a list of candidate models. Below we denote the i^{th} model as $\mathcal{M}_i$ which has a parameter vector $\tilde{\boldsymbol{\theta}}_i$ for $i = 1, \ldots, K$ where K is the number of candidate models being considered.

Here we explain the formulation of the RJMCMC algorithm developed by Barker and Link (2013). They relate the parameters of all candidate models using a palette of parameters which is denoted as $\tilde{\boldsymbol{\psi}}$. The parameter vector for the i^{th} model is obtained using a bijective function

$$g_i(\tilde{\boldsymbol{\psi}}_i) = \boldsymbol{\xi}_i = \begin{bmatrix} \tilde{\boldsymbol{\theta}}_i \\ \boldsymbol{u}_i \end{bmatrix}$$

where $\boldsymbol{u}_i$ is a random parameter vector introduced such that $\dim(\tilde{\boldsymbol{\psi}}) = \dim(\boldsymbol{\xi}_i)$. The $\boldsymbol{u}_i$ variables are introduced so as to ensure that the Gibbs algorithm developed below maintains 'detailed balance'. Below we specifically assume that $g_i()$ is an identity function $(g_i(\tilde{\boldsymbol{\psi}}_i) = \tilde{\boldsymbol{\psi}}_i)$ which aids in the calculation of equation 5.14 below.

Link and Barkers' method is particularly convenient since it allows one to perform RJMCMC in two stages. In the first stage, all candidate models are independently fit to data and the posterior MCMC samples of the parameters of the candidate models are stored. In the second step, we use these samples to calculate the posterior probabilities of each candidate model. The algorithm entails sampling from the following conditional distributions in turn,

$$\begin{aligned} &[\mathcal{M}|\boldsymbol{y}, \tilde{\boldsymbol{\psi}}] \text{ for } \mathcal{M} = M_1, \ldots, M_K, \\ &[\tilde{\boldsymbol{\psi}}|\boldsymbol{y}, \mathcal{M}] \end{aligned} \tag{5.13}$$

and in so doing, obtains the marginal posterior probabilities of $\Pr(\mathcal{M}_i|\boldsymbol{y})$ for $i = 1, \ldots, K$.

$[\mathcal{M}|\boldsymbol{y}, \tilde{\boldsymbol{\psi}}]$ is a categorical distribution with

$$\Pr(\mathcal{M}_i|.) = \frac{[\boldsymbol{y}|\tilde{\boldsymbol{\psi}}, \mathcal{M}_i][\tilde{\boldsymbol{\psi}}|\mathcal{M}_i][\mathcal{M}_i]}{\sum_{j=1}^{K}[\boldsymbol{y}|\tilde{\boldsymbol{\psi}}, \mathcal{M}_j][\tilde{\boldsymbol{\psi}}|\mathcal{M}_j][\mathcal{M}_j]}. \tag{5.14}$$

Assuming $\tilde{\boldsymbol{\theta}}_i$ and $\boldsymbol{u}_i$ are independent, $[\tilde{\boldsymbol{\psi}}|\mathcal{M}_i] = [\tilde{\boldsymbol{\theta}}_i|\mathcal{M}_i][\boldsymbol{u}_i|\mathcal{M}_i]$. $[\tilde{\boldsymbol{\theta}}_i|\mathcal{M}_i]$ represents the distribution of the parameters of model $\mathcal{M}_i$ while $[\boldsymbol{u}_i|\mathcal{M}_i]$ is the joint prior density of the vector $\boldsymbol{u}_i$. In Section 5.5 we assume that all u_i are independent and identically distributed standard Gaussian random variables. $[\boldsymbol{y}|\tilde{\boldsymbol{\psi}}, \mathcal{M}_i]$ represents the joint density of the data for model i (refer to equation (5.12)). For moderate sized sized values of K, sampling from $[\mathcal{M}|\boldsymbol{y}, \tilde{\boldsymbol{\psi}}]$ is straightforward. Posterior draws from $[\tilde{\boldsymbol{\psi}}|\boldsymbol{y}, \mathcal{M}]$ can also easily be obtained by noting that $\tilde{\boldsymbol{\psi}}_i = g_i^{-1}(\boldsymbol{\xi}_i)$. A random draw of $\boldsymbol{\xi}_i$ is obtained by randomly drawing from the MCMC samples stored previously as well as sampling from $[\boldsymbol{u}_i|\mathcal{M}_i]$.

With regards to the model used in Section 5.5, the denominator in equation (5.14) evaluated

at the k^{th} iteration of equations (5.13) simplifies to

$$\begin{aligned}\Pr(\mathcal{M}_i|.) \propto & [\boldsymbol{y}|\boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{z}^{(k)}][\boldsymbol{z}^{(k)}|\boldsymbol{\beta}^{(k)}] \\ & \times [\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{(k)}] \prod_{i=1}^{n_s} [\boldsymbol{\alpha}_i^{(k)}|\boldsymbol{\mu}_{\boldsymbol{\alpha}}^{(k)}, \sigma_{\boldsymbol{\alpha}}^{2(k)}] \\ & \times [\boldsymbol{\mu}_{\boldsymbol{\beta}}^{(k)}] \prod_{i=1}^{n_s} [\boldsymbol{\beta}_i^{(k)}|\boldsymbol{\mu}_{\boldsymbol{\beta}}^{(k)}, \sigma_{\boldsymbol{\beta}}^{2(k)}] \\ & \times [\boldsymbol{u}^{(k)}]\end{aligned} \tag{5.15}$$

where the dependence on $\mathcal{M}_i$ in the above equation has been omitted for notational convenience and $[\boldsymbol{u}]$ is appropriately dimensioned.

Barker and Link (2013) suggested an alternate method of obtaining the posterior model probabilities. This entails estimating the transition probability matrix $\boldsymbol{T}$ and then normalising the left eigenvector of $\boldsymbol{T}$ where the $(i, j)^{\text{th}}$ entry of $\boldsymbol{T}$ contains the probability of 'moving' from model i to model j (Seber, 2008). This method was undertaken in Section 5.5.

5.H Model selection and goodness-of-fit statistics

Table 21. Model selection and goodness-of-fit statistics for each model using ten independent MCMC runs.

Model	Bayesian p-value			Sampled Bayesian p-value			WAIC		
	Mean	SE	Median	Mean	SE	Median	Mean	SE	Median
1	0.29	0.002	0.29	0.28	0.04	0.29	3662.60	0.35	3662.39
2	0.26	0.001	0.26	0.27	0.04	0.25	3813.74	0.50	3813.70
3	0.14	0.001	0.14	0.17	0.06	0.11	4064.99	0.63	4064.21
4	0.19	0.002	0.19	0.16	0.04	0.16	4134.79	0.53	4135.09
5	0.62	0.001	0.61	0.57	0.06	0.63	3666.69	0.33	3666.86
6	0.55	0.003	0.55	0.49	0.06	0.47	3787.54	0.37	3787.54
7	0.28	0.002	0.28	0.25	0.07	0.19	3871.90	0.39	3871.60
8	0.32	0.002	0.32	0.47	0.08	0.42	3963.64	0.51	3964.32
9	0.57	0.002	0.57	0.57	0.06	0.54	3628.49	0.37	3628.65
10	0.48	0.002	0.48	0.49	0.06	0.50	3764.34	0.44	3763.82
11	0.26	0.002	0.26	0.32	0.07	0.28	3970.58	0.58	3970.80
12	0.27	0.002	0.27	0.32	0.07	0.27	4046.11	0.59	4045.73
13	0.72	0.002	0.72	0.77	0.03	0.78	3646.08	0.33	3646.18
14	0.64	0.002	0.64	0.77	0.05	0.82	3761.54	0.37	3761.76
15	0.39	0.003	0.39	0.32	0.04	0.33	3857.95	0.77	3859.34
16	0.40	0.002	0.41	0.46	0.05	0.46	3942.09	0.58	3942.38
17	0.70	0.002	0.70	0.68	0.04	0.66	3778.67	0.57	3778.47
18	0.69	0.002	0.69	0.76	0.06	0.85	3928.14	0.85	3928.85
19	0.32	0.002	0.32	0.22	0.04	0.24	4010.34	0.76	4010.08
20	0.40	0.002	0.40	0.40	0.09	0.32	4096.97	0.82	4096.75
21	0.82	0.001	0.81	0.84	0.03	0.83	3725.81	0.41	3726.04
22	0.79	0.002	0.79	0.85	0.04	0.89	3844.02	0.51	3844.00
23	0.50	0.002	0.50	0.57	0.05	0.54	3887.98	0.78	3887.80
24	0.57	0.003	0.57	0.59	0.07	0.56	3986.70	0.88	3986.66
25	0.86	0.001	0.86	0.89	0.02	0.89	3736.57	0.44	3736.55
26	0.82	0.001	0.82	0.85	0.03	0.86	3864.17	0.51	3864.22
27	0.57	0.002	0.57	0.60	0.07	0.64	3959.82	0.97	3960.46
28	0.59	0.002	0.59	0.61	0.07	0.67	4036.85	0.67	4037.05
29	0.89	0.001	0.89	0.87	0.03	0.89	3710.03	0.29	3709.92
30	0.86	0.001	0.86	0.87	0.03	0.88	3821.69	0.56	3822.19
31	0.64	0.002	0.64	0.72	0.05	0.73	3883.59	1.23	3884.72
32	0.68	0.002	0.68	0.68	0.07	0.68	3973.62	0.48	3973.24

5.I Regression effects for models proposed

Table 22. The community-level regression effects of Models 5, 6, 8, 22 and 24. (Posterior mean, Standard deviation, 2.5% and 97.5% quantiles.) Coefficients where zero does not lie in the 95% equal-tail credibility interval are bolded.

Model	Coefficient	Mean	Std	2.5%	97.5%
6	μ_{α_0}	**-1.88**	0.19	-2.28	-1.51
	$\mu_{\alpha_{\text{trail}}}$	**-0.55**	0.21	-0.97	-0.14
	μ_{β_0}	**-1.90**	0.23	-2.35	-1.46
	$\mu_{\beta_{\text{elevation}}}$	**-0.59**	0.22	-1.03	-0.17
8	μ_{α_0}	**-2.44**	0.24	-2.93	-1.99
	$\mu_{\alpha_{\text{trail}}}$	**-0.50**	0.22	-0.93	-0.08
	$\mu_{\alpha_{\text{plain}}}$	0.13	0.24	-0.34	0.58
	$\mu_{\alpha_{\text{river}}}$	0.16	0.25	-0.34	0.65
	μ_{β_0}	**-1.63**	0.24	-2.12	-1.17
	$\mu_{\beta_{\text{elevation}}}$	**-0.57**	0.22	-1.02	-0.13
22	μ_{α_0}	**-1.92**	0.21	-2.34	-1.54
	$\mu_{\alpha_{\text{trail}}}$	**-0.57**	0.22	-1.00	-0.16
	μ_{β_0}	**-1.94**	0.23	-2.39	-1.47
	$\mu_{\beta_{\text{elevation}}}$	**-0.58**	0.22	-1.02	-0.16
	$\mu_{\beta_{\text{PreyIndex}}}$	0.20	0.23	-0.24	0.64
24	μ_{α_0}	**-2.39**	0.23	-2.87	-1.94
	$\mu_{\alpha_{\text{trail}}}$	**-0.53**	0.22	-0.96	-0.11
	$\mu_{\alpha_{\text{plain}}}$	0.06	0.23	-0.41	0.50
	$\mu_{\alpha_{\text{river}}}$	0.05	0.25	-0.44	0.54
	μ_{β_0}	**-1.69**	0.24	-2.17	-1.24
	$\mu_{\beta_{\text{elevation}}}$	**-0.58**	0.23	-1.03	-0.14
	$\mu_{\beta_{\text{PreyIndex}}}$	0.25	0.23	-0.20	0.70
5	μ_{α_0}	**-1.88**	0.17	-2.23	-1.56
	μ_{β_0}	**-2.17**	0.22	-2.63	-1.74
	$\mu_{\beta_{\text{elevation}}}$	**-0.66**	0.22	-1.09	-0.24

5.J Occupancy probability map for all species

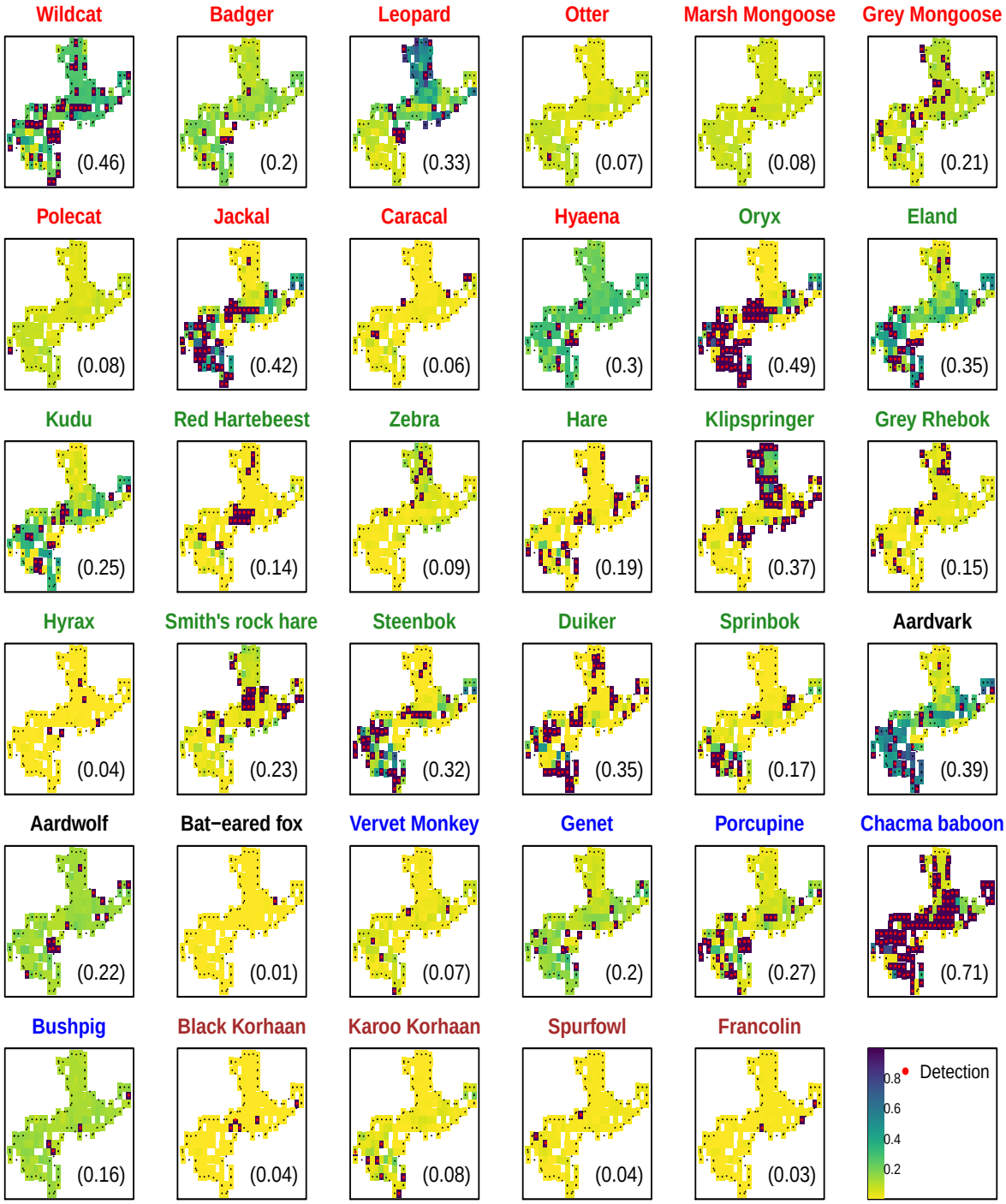

Figure 30. Estimated mean occupancy probabilities for each of the species investigated in the Anysberg Nature Reserve (longitude and latitude is not included in the plots). The mean occupancy probability for each species over the region is displayed in brackets. The results were grouped into the five groups namely carnivores (red), herbivores (green), insectivores (black), omnivores (blue) and birds (brown).

5.K Statistics pertaining to Model 6

Below we display some statistics pertaining to the community-level regression effects for Model 6 when run using three chains of 100 000 in length. A burn-in sample of one-third is specified thereafter the resulting chains are thinned by retaining every 10^{th} observation. In order to investigate the sensitivity of the results to the prior distribution, the results from four different prior distributions are displayed below. The different prior distributions are specified as: Prior 1: $a^2 = 2.25^2 = b^2$, $A^2 = 2.25^2 = B^2$; Prior 2: $a^2 = 2.25^2 = b^2$, $A^2 = 2.25 = B^2$; Prior 3: $a^2 = 5^2 = b^2$, $A^2 = 5^2 = B^2$; Prior 4: $a^2 = 5^2 = b^2$, $A^2 = 5 = B^2$.

From Table 23 and Figure 31 we see that the posterior statistics for Model 6 is very similar for all four prior distributions specified. The results for the remaining parameters of the model are not displayed but it was found that the resulting posterior distributions were not sensitive to the particular prior distribution specified.

Table 23. The community-level regression effects of Model 6 for four different prior distribution specifications. (Posterior mean, Standard deviation, Naive standard error and time-series standard error as calculated by the *coda* package.)

Prior	Coefficients	**M**ean	**S**td	**Naive SE**	**Time-series SE**
1	μ_{α_0}	-1.88	0.19	0.0014	0.0018
	$\mu_{\alpha_{\text{trail}}}$	-0.55	0.21	0.0015	0.0018
	μ_{β_0}	-1.90	0.23	0.0016	0.0023
	$\mu_{\beta_{\text{elevation}}}$	-0.59	0.22	0.0016	0.0017
2	μ_{α_0}	-1.88	0.19	0.0014	0.0016
	$\mu_{\alpha_{\text{trail}}}$	-0.55	0.21	0.0015	0.0018
	μ_{β_0}	-1.91	0.23	0.0016	0.0024
	$\mu_{\beta_{\text{elevation}}}$	-0.59	0.22	0.0015	0.0017
3	μ_{α_0}	-1.89	0.20	0.0014	0.0018
	$\mu_{\alpha_{\text{trail}}}$	-0.55	0.21	0.0015	0.0019
	μ_{β_0}	-1.92	0.23	0.0016	0.0025
	$\mu_{\beta_{\text{elevation}}}$	-0.60	0.22	0.0015	0.0018
4	μ_{α_0}	-1.89	0.20	0.0014	0.0020
	$\mu_{\alpha_{\text{trail}}}$	-0.55	0.21	0.0015	0.0017
	μ_{β_0}	-1.92	0.23	0.0016	0.0027
	$\mu_{\beta_{\text{elevation}}}$	-0.60	0.22	0.0015	0.0017

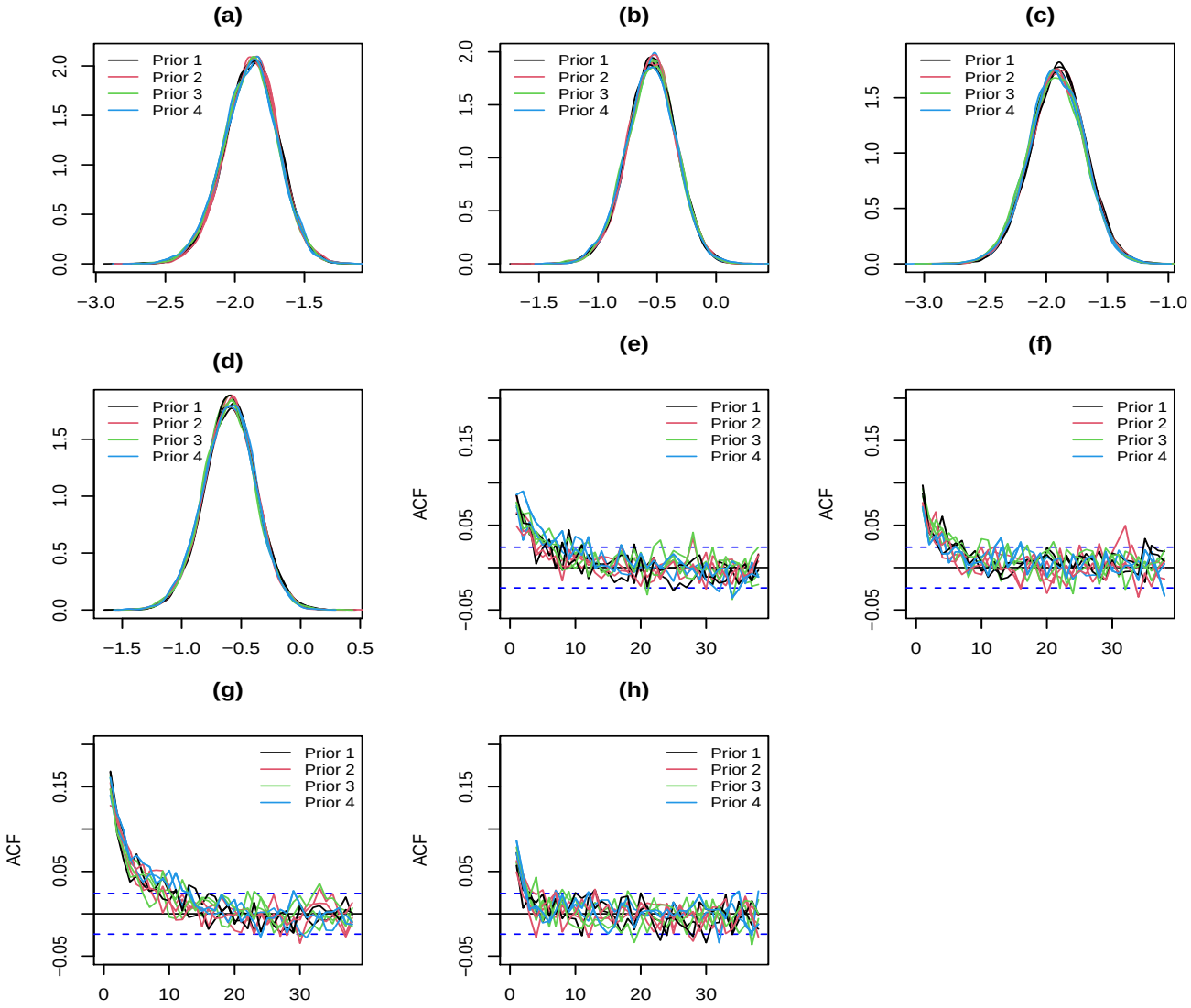

Figure 31. Posterior distributions ((a)= μ_{α_0}, (b)= $\mu_{\alpha_{\text{trail}}}$, (c)= μ_{β_0}, (d)= $\mu_{\beta_{\text{elevation}}}$) and estimated lagged sample autocorrelations ((e)= μ_{α_0}, (f)= $\mu_{\alpha_{\text{trail}}}$, (g)= μ_{β_0}, (h)= $\mu_{\beta_{\text{elevation}}}$) of the posterior samples of the community-level regression effects for Model 6 when run using three chains of 100 000 in length. A burn-in sample of one-third is specified thereafter the resulting chains are thinned by retaining every 10^{th} observation. The results of four different prior distributions are displayed. We see that the resulting posterior distributions are not sensitive to the prior distribution used and that the level of autocorrelation among the chains used for all prior distributions are very similar.

5.L Comparison between the Gibbs sampler developed and *Stan*

Below we display some statistics pertaining to the community-level regression effects for Model 6 when run using three chains of 100 000 in length. A burn-in sample of one-third is specified thereafter the resulting chains are thinned by retaining every 10^{th} observation. We utilize the effective sample size (ESS) and the effective sampling rate (ESR) (effective sample size per second of run-time) to compare the efficiency of MCMC chains produced using the Gibbs sampling algorithm developed here with that of an implementation of the model using *Stan* (Stan Development Team, 2020b). Prior 1 described in Appendix 5.K is used here.

From Figure 32 we see that both MCMC methods produce the same posterior distribution of the regression effects of the community-level regression effects although the Gibbs sampler produces significantly larger positive autocorrelations in the simulated chains at small lags compared to the those generated when using *Stan*.

Table 24. Effective sample size obtained using *Stan*

	Gibbs	*Stan*
μ_{α_0}	12 374.79	19 510.64
$\mu_{\beta_{\text{elevation}}}$	13 091.33	19 422.46
μ_{β_0}	9 082.96	20 007.89
$\mu_{\beta_{\text{elevation}}}$	14 989.48	18 897.94

Table 24 displays the effective sample size obtained for the two MCMC algorithms. The R package *rstan* (Stan Development Team (2020a)) is used to perform these calculations. Table 25 displays the ratio of the ESR for the two MCMC methods (i.e. $\text{ESR}_{Gibbs}/\text{ESR}_{Stan}$). Three different R packages are used to undertake the calculation of the effective sample sizes since there has been much debate regarding which method is most appropriate in the literature (Plummer et al., 2007; Flegal et al., 2020; Stan Development Team, 2020a). We see that in general *Stan* generates larger effective sample sizes for all of the parameters considered here. The Gibbs algorithm took approximately 13 minutes to complete the

sampling process while *Stan* took approximately 4.6 hours. This implies that the Gibbs algorithm has an ESR approximately $9 - 17$ times larger than the ESR of *Stan*.

Table 25. The ratio of the effective sampling rates of the two MCMC methods using three different R packages to calculate the effective sample size.

Coefficient	*coda*	*mcmcse*	*rstan*
μ_{α_0}	13.11	14.61	13.63
$\mu_{\beta_{\text{elevation}}}$	14.85	12.78	14.49
μ_{β_0}	10.32	9.42	9.76
$\mu_{\beta_{\text{elevation}}}$	17.76	14.81	17.05

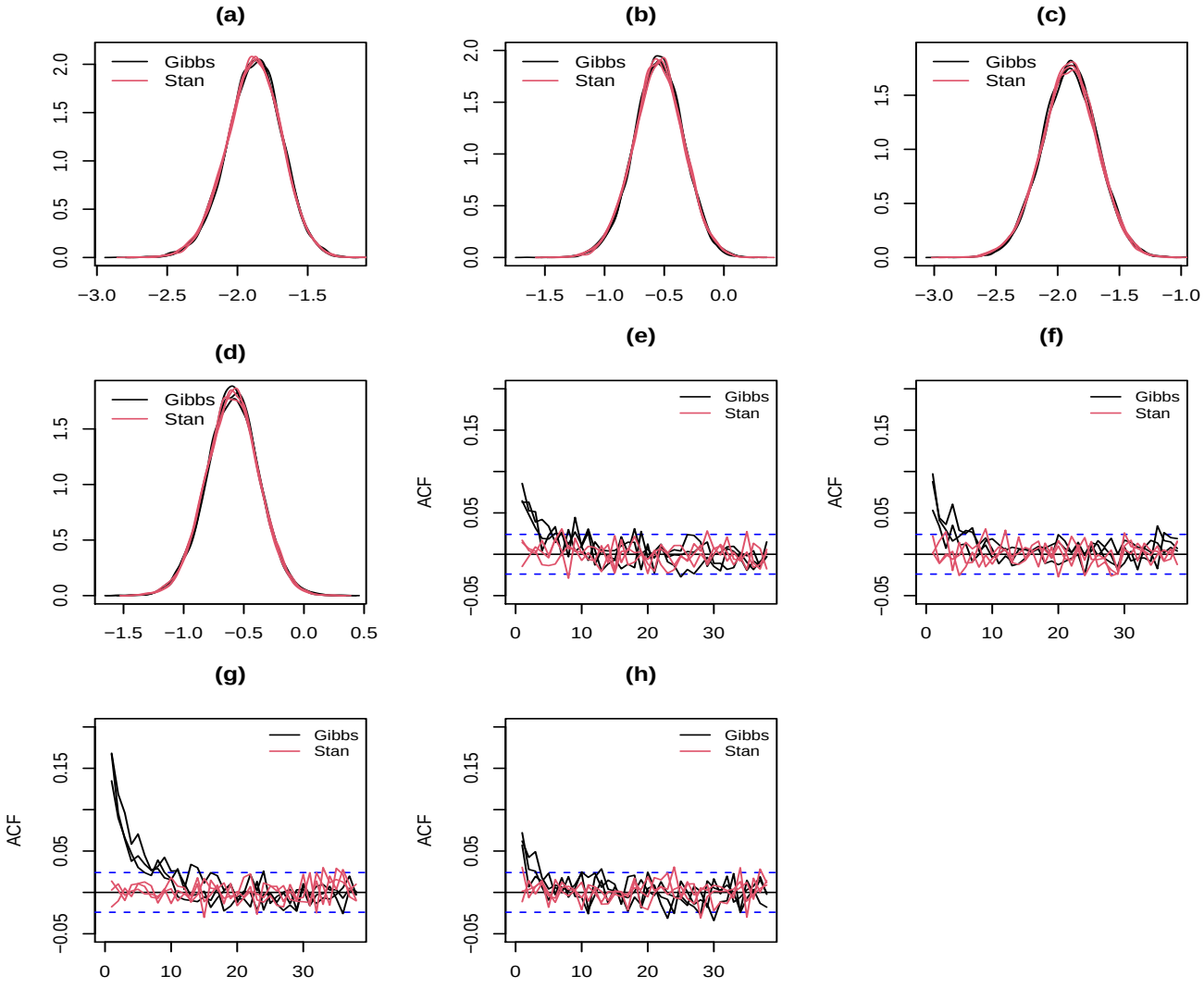

Figure 32. Posterior distributions ((a)= μ_{α_0}, (b)= $\mu_{\alpha_{\text{trail}}}$, (c)= μ_{β_0}, (d)= $\mu_{\beta_{\text{elevation}}}$) and estimated lagged sample autocorrelations ((e)= μ_{α_0}, (f)= $\mu_{\alpha_{\text{trail}}}$, (g)= μ_{β_0}, (h)= $\mu_{\beta_{\text{elevation}}}$) of the posterior samples of the community-level regression effects for Model 6 when run using three chains of 100 000 in length. A burn-in sample of one-third is specified thereafter the resulting chains are thinned by retaining every 10^{th} observation. The Gibbs algorithm and *Stan* are compared.

The *Stan* code

The *Stan* code for the above MSO model is included below.

```
stancode <- "
//See https://mc-stan.org/users/documentation/case-studies
/dorazio-royle-occupancy.html

functions {
   real lp_observed(int x, int K, real logistic_psi, real logistic_p)
   {
      return log_inv_logistic(logistic_psi) +
      binomial_logistic_lpmf(x | K, logistic_p);
   }

   real lp_unobserved(int K, real logistic_psi, real logistic_p)
   {
      return log_sum_exp(lp_observed(0, K, logistic_psi, logistic_p),
      log1m_inv_logistic(logistic_psi));
   }
}

data {
   //prior parameters
   real<lower=0> a;
   real<lower=0> b;
   real<lower=0> A2;
   real<lower=0> B2;

   int<lower=1> ns; //number of species
   int<lower=1> nsites;
   int<lower=1> no; //number of occupancy covariates
   int<lower=1> nd; //number of detection covariates

   int V[nsites]; //number of visits to each site

   //note the 17 is hard coded. This happens to be maxima
   // for the data analysed
   int<lower=0, upper= 17> y[ns, nsites]; //Binomial data
   matrix[nsites, no] Xmat; //occupancy covariates
   matrix[nsites, nd] Dmat; //detection covariates
}

parameters {
   matrix[nd, ns] alpha; //alpha coeffs - detection regression effects
   matrix[no, ns] beta; //beta coeffs - occupancy regression effects

   vector[nd] mualpha; //community detection effects
   vector[no] mubeta; // community occuoancy effects

   real<lower=0> s2alpha;
   real<lower=0> nabla_alpha;
   real<lower=0> s2beta;
   real<lower=0> nabla_beta;
}

transformed parameters {
```

```
   //note the dimension here!
   matrix[nsites, ns] logistic_psi;    // log odds  of occurrence
   matrix[nsites, ns] logistic_p;  // log odds of detection

   logistic_psi = Xmat*beta;
   logistic_p = Dmat*alpha;
}

model {
   //priors
   nabla_alpha ~ inv_gamma(0.5, 1/A2);
   s2alpha ~ inv_gamma(0.5, 1/pow(nabla_alpha,2));

   nabla_beta ~ inv_gamma(0.5, 1/B2);
   s2beta ~ inv_gamma(0.5, 1/pow(nabla_beta,2));

   //mualpha ~ multi_normal(rep_vector(0, nd), identity_matrix(nd)*a2);
   //mubeta ~ multi_normal(rep_vector(0, no), identity_matrix(no)*b2);

   for (i in 1:nd){ mualpha[i] ~ normal(0, a);}

      for (i in 1:no){ mubeta[i] ~ normal(0, b);}

      /*
      for (ispecies in 1:ns){
         alpha[, ispecies] ~ multi_normal(mualpha, identity_matrix(nd)*
   s2alpha);
      }*/

   //prior distribution for elements of alpha
   for (ispecies in 1:ns){
      for (idet in 1:nd){
         alpha[idet, ispecies] ~ normal(mualpha[idet], sqrt(s2alpha));
      }
   }

   //prior distribution for elements of beta
   for (ispecies in 1:ns){
      for (iocc in 1:no){
         beta[iocc, ispecies] ~ normal(mubeta[iocc], sqrt(s2beta));
      }
   }

   // likelihood
   for (i in 1:ns) {
      for (j in 1:nsites) {
         if (y[i,j] > 0)
         //notice the indices for logistic_psi and logistic_p
         target += lp_observed(y[i,j], V[j],
         logistic_psi[j,i], logistic_p[j,i]);
      else
      target += lp_unobserved(V[j], logistic_psi[j,i], logistic_p[j,i]);
      }
   }//end likelihood
}
"
```

Chapter 6

Conclusions

6.1 Introduction

The main aim of this thesis was to develop and assess the effectiveness of statistical methods that can be used to fit various types of occupancy models to detection-nondetection data. The three main areas investigated were: (1) Variational Bayes methods applied to single season occupancy models; (2) the development of a fast and efficient Gibbs algorithm for fitting Bayesian nonspatial and spatial single season occupancy models when the regression effects of the occupancy model are modelled using logistic link functions and (3) the development of various Gibbs algorithms for fitting multi-species occupancy models. A secondary aim of the study was to develop a publicly available R package which allowed users to implement the methods developed here on their data.

Below, I provide a summary of the key findings of the thesis as well as some areas of future development.

6.2 A summary of the key findings of the thesis

In Chapter Two and Chapter Three, I Variational Bayes approximations to the posterior distributions of the parameters of the single season (single species) occupancy model that uses *logistic* and *probit* link functions respectively. In Chapter Two, two new methods were developed for the *logistic* link function case. The first method uses a Laplace approximation of the Variational Bayesian optimal distributions while the second method utilises the tangent based method of Jaakkola and Jordan (2000) to obtain the Variational Bayes approximations. These methods were employed to approximate the predictive distribution of the proportion of occupied sites in a region. Based on a simulation study I found that the Laplace approximation method performed well under most conditions considered and was suitable in approximating the posterior distributions of the parameters of the single season (single species) occupancy model while the *tangent based method* performed poorly. In general, the Laplace method attains very similar frequentist coverage probabilities to those obtained by the maximum likelihood estimation method and it is advised that the method could be used when the detection probability is at least 0.5 and there are at least three sampling occasions.

Secondly, simulation results showed that the Laplace approximation method provide reasonable approximate distributions of the predictive distribution of the proportion of area occupied. In scenarios where the detection probabilities are relatively low and the number of survey occasions are small (at most two), the Laplace method slightly underestimated the upper bounds of the predictive distribution of the proportion of area occupied although the differences between the true predictive distribution and the approximate one were small when at least three survey occasions are undertaken.

In Chapter Three, two Variational Bayes methods focusing specifically on the *probit* link function case were developed. The first of these methods uses multivariate Gaussian distributions as the approximating distributions, whereas the second method used linear Gaussian mixture distributions. The methods were employed to approximate the predictive distribution of the proportion of occupied sites in a region. Based on a simulation study, it was shown that Gaussian mixture distributions produced better approximations to the marginal posterior distributions of the regression effects of the model than when multivariate Gaussian distributions were used. It was also found that the accuracy of the approximations to the marginal distribution of the regression effects of the model increased as the number of sites visited and the number of survey occasions increased. Similar results were found when approximating the predictive distribution of the proportion of area occupied.

In Chapter Four, the link between *logistic* regression and occupancy models was exploited to develop Gibbs sampling algorithms for (nonspatial and spatial single species) occupancy models when the regression effects of the of the models are modelled using *logistic* link functions. Two Gibbs sampling algorithms were developed. The first uses the *difference of random utility method* (Algorithm 1) of Frühwirth-Schnatter and Frühwirth (2010) while the second uses Pólya-Gamma sampling (Algorithm 2) and data augmentation (Polson et al., 2013). The expected sample size and expected sampling rate were used to compare the two methods developed to the *JAGS* and *Stan* implementations of the single season (single species) nonspatial occupancy model via a simulation study which focused on evaluating the run-times of the above algorithms.

It was concluded that Algorithm 2 had the fastest **run-times** across all scenarios considered. *JAGS* and *Stan* had similar run-times when the number of sites included in the study was no more than one hundred attaining median run-times being between $6-12$ times slower than Algorithm 2. The *JAGS* algorithm did not scale well and took $40-70$ times longer to complete relative to Algorithm 2 when five hundred sites were included in the study. It was found that the expected sample size obtained using *Stan* was significantly larger than those obtained when using the other algorithms. *JAGS* and Algorithm 1 produced similar expected sample sizes while Algorithm 2 produced larger expected sample sizes

than *JAGS* and Algorithm 1 in general.

The median expected sampling rate obtained using Algorithm 2 was **generally much larger** than the median expected sampling rate when using the other three algorithms and occurred largely due to the faster run-times of Algorithm 2. When the number of sites visited was limited to one hundred, the expected sampling rate for Algorithm 2 was $3-12$ times larger than those obtained using *Stan*; $3-5$ times larger than those obtained using Algorithm 1 and $10-120$ times larger than those obtained when using *JAGS*. When five hundred sites are included in the study, Algorithm 2 performs significantly better than all other algorithms with the expected sampling rate being $3-10$ times larger than those obtained when using *Stan*, $15-60$ times larger than those obtained when using Algorithm 1 and $30-450$ times larger than those obtained when using *JAGS*.

An R package has been developed to fit the models developed in Chapters Two-Four and is freely available at `https://github.com/AllanClark/Rcppocc`.

Two types of multi-species occupancy models exist in the literature. The first model-type is used to estimate species richness (Dorazio and Royle, 2005) in an area while the second model-type is used to obtain species specific regression effects when the species richness is known (Broms et al., 2016). Several studies have been undertaken using both model-types. These studies however either use probit link functions to obtain the regression effects of their models; use general Bayesian analysis software such as *JAGS*, *WinBUGS*, *NIMBLE* or *Stan* to undertake their analysis or use the Metropolis Hastings algorithm to sample from the parameters of their models. Gibbs sampling algorithms that can be used when *probit* or *logistic* link functions are employed to model the detection and occupancy processes respectively were developed in Chapter Five. The Chapter also shows how this can be done for a number of different versions of the multi-species occupancy model. It specifically highlights that the algorithms for multi-species occupancy models (that use a *logistic* link function) are similar to those developed for single season occupancy models as well as single season spatial occupancy models where all algorithms use the work of Polson et al. (2013) and Clark and Altwegg (2019) to develop efficient Gibbs sampling algorithms.

An R package has been developed to fit some of the models developed in Chapter Five and is freely available at `https://github.com/AllanClark/MSO`.

6.3 Some developments since 2019

Two methodological advancements that references previously published parts of this thesis, are discussed below, namely Diana et al. (2021) and Doser et al. (2021). Both papers assume that the regression effects of the detection and occupancy processes of their models

are modelled using logistic link functions and utilises the Pólya-Gamma formulation described by Clark and Altwegg (2019) in their algorithms.

Diana et al. (2021) developed a model that allows one to estimate the occupancy and detection probability estimates of a region for many years. They do this by modelling the evolution of the occupancy and detection probability through time using Gaussian processes (Williams and Rasmussen, 2006).

Doser et al. (2021) extended Clark and Altwegg (2019) and developed algorithms to fit various Bayesian occupancy models. As one of their developments, they assume that the spatial effects in the occupancy process of a single-season, single-species occupancy model can be modelled through a Gaussian process prior distribution with a Matérn covariance function. They found that for data sets with more than 1 000 sites, their algorithm was prohibitively slow in which case they used a nearest-neighbourhood Gaussian process prior distribution (Datta et al., 2016) to significantly reduce run-times when analysing large detection-nondetection data sets.

6.4 Areas of future development

There are several areas of this thesis that could be developed further.

The Variational Bayes methods for the single season (single species) occupancy model developed in Chapters Two and Three do not include random effects in the detection and occupancy processes. One reason for not including random effects in the detection process is because one does not expect there to be significant spatial autocorrelation present in the detection process of SABAP2 data. Considering this has been required for some previous studies (Wright et al., 2019), future studies may test for spatial autocorrelation in the detection process and expand the models developed in this thesis to account for spatial autocorrelation. Importantly, Variational Bayes methods for the spatial models developed in Chapter Four or the multi-species models found in Chapter Five were not developed in this thesis since it found that at times the Variational Bayes methods slightly underestimated the upper bounds of the predictive distribution of the proportion of area occupied as well as the regression effects of the single season occupancy model. In the future, it might be useful to explore these developments since the spatial and multi-species occupancy models require significant run-times to complete in general and the researcher could make a trade-off between computational time and underestimated posterior variances of the parameters of their model themselves.

In Chapter Four, MCMC algorithms for a single season (single species) nonspatial and spatial occupancy models were developed. The model could be extended to include random effects in the detection process and a Gibbs sampler could easily be developed for the

integrated conditional autoregressive model discussed in Johnson et al. (2013). In this chapter spatial autocorrelation is modelled by using restricted spatial regression. In future different spatial models could be used to model the spatial structure that might exist in the occupancy process. In Chapter Four (published as Clark and Altwegg (2019)) the efficiency of the Gibbs sampling algorithms developed for the single season (single species) occupancy models are compared them when implementing the model in *JAGS* and *Stan* although no comparison was made to a *NIMBLE* implementation of the model.

In Chapter Five, Gibbs algorithms for various multi-species occupancy models were developed and R code (R Core Team, 2014) for a small subset of the models was provided. There is thus a need to develop a comprehensive R repository that incorporates all of the models discussed in this thesis.

Bibliography

Aing, C., Halls, S., Oken, K., Dobrow, R., and Fieberg, J. (2011). A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites. *Journal of Applied Ecology*, 48(6):1508–1517.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Altwegg, R., Wheeler, M., and Erni, B. (2008). Climate and the range dynamics of species with imperfect detection. *Biology letters*, 4(5):581–584.

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association*, 49(268):765–769.

Andrews, D. F. and Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(1):99–102.

Animal Demography Unit (2014). The Southern African Bird Atlas Project 2. `http://sabap2.adu.org.za/`. [Online; accessed November 25, 2014].

Baddeley, A. and Turner, R. (2005). Spatstat: An R package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12(6):1–42.

Bailey, L. L., MacKenzie, D. I., and Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, 5(12):1269–1279.

Balakrishnan, N. (1991). *Handbook of the logistic distribution*. CRC Press.

Barker, R. J. and Link, W. A. (2013). Bayesian multimodel inference by RJMCMC: A Gibbs sampling approach. *The American Statistician*, 67(3):150–156.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science and Business Media.

Bernardo, J. M. and Smith, A. F. (1994). *Bayesian theory*, volume 405. John Wiley and Sons.

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(4):691–746.

Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82(4):733–746.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Bishop, C. M. (2008). A New Framework for Machine Learning. In Zurada, J. M., Yen, G. G., and Wang, J., editors, *Computational Intelligence: Research Frontiers: IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, China, June 1-6, 2008, Plenary/Invited Lectures*, pages 1–24. Springer Berlin Heidelberg.

Bled, F., Nichols, J. D., and Altwegg, R. (2013). Dynamic occupancy models for analyzing species' range dynamics across large geographic scales. *Ecology and Evolution*, 3(15):4896–4909.

Boehm, L., Reich, B. J., and Bandyopadhyay, D. (2013). Bridging conditional and marginal inference for spatially referenced binary data. *Biometrics*, 69(2):545–554.

Bornkamp, B. (2011). Approximating probability densities by iterated Laplace approximations. *Journal of Computational and Graphical Statistics*, 20(3):656–669.

Bornkamp, B. and Bornkamp, M. B. (2012). Package 'iterlap'. *Journal of Computational and Graphical Statistics*, 20(3):656–669.

Broms, K. M. (2013). *Using Presence-Absence Data on Areal Units to Model the Ranges and Range Shifts of Select South African Bird Species*. PhD thesis, University of Washington.

Broms, K. M., Hooten, M. B., and Fitzpatrick, R. M. (2016). Model selection and assessment for multi-species occupancy models. *Ecology*, 97(7):1759–1770.

Broms, K. M., Johnson, D. S., Altwegg, R., and Conquest, L. L. (2014). Spatial occupancy models applied to atlas data show Southern Ground Hornbills strongly depend on protected areas. *Ecological Applications*, 24(2):363–374.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC handbooks of modern statistical methods. Chapman and Hall/CRC.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32.

Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, 1(4):651–673.

Choi, H. M. and Hobert, J. P. (2013). The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electronic Journal of Statistics*, 7:2054–2064.

Choy, S. L., O'Leary, R., and Mengersen, K. (2009). Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. *Ecology*, 90(1):265–277.

Clark, A. E. and Altwegg, R. (2019). Efficient Bayesian analysis of occupancy models with logit link functions. *Ecology and evolution*, 9(2):756–768.

Clark, A. E., Altwegg, R., and Ormerod, J. T. (2016). A variational Bayes approach to the analysis of occupancy models. *PLoS ONE*, 11(2):e0148966.

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology letters*, 8(1):2–14.

Conn, P. B., Johnson, D. S., Williams, P. J., Melin, S. R., and Hooten, M. B. (2018). A guide to Bayesian model checking for ecologists. *Ecological Monographs*, 88(4):526–542.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and 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.

Devarajan, K., Morelli, T. L., and Tenan, S. (2020). Multi-species occupancy models: review, roadmap, and recommendations. *Ecography*, 43(11):1612–1624.

Diana, A., Dennis, E., Matechou, E., and Morgan, B. (2021). Fast bayesian inference for large occupancy data sets, using the polya-gamma scheme. *arXiv preprint arXiv:2107.14656*.

Dorazio, R. M., Gotelli, N. J., and Ellison, A. M. (2011). Modern methods of estimating biodiversity from presence-absence surveys. In *Biodiversity Loss in a Changing Planet.*, pages 277–302. InTech Rijeka, Croatia.

Dorazio, R. M. and Rodriguez, D. T. (2012). A Gibbs sampler for Bayesian analysis of site-occupancy data. *Methods in Ecology and Evolution*, 3(6):1093–1098.

Dorazio, R. M. and Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, 100(470):389–398.

Dorazio, R. M., Royle, J. A., Söderström, B., and Glimskär, A. (2006). Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, 87(4):842–854.

Doser, J. W., Finley, A. O., Kéry, M., and Zipkin, E. F. (2021). spoccupancy: An r package for single species, multispecies, and integrated spatial occupancy models. *arXiv preprint arXiv:2111.12163*.

Drouilly, M., Clark, A., and O'Riain, M. J. (2018). Multi-species occupancy modelling of mammal and ground bird communities in rangeland in the karoo: A case for dryland systems globally. *Biological Conservation*, 224:16–25.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics.

Fiske, I. and Chandler, R. (2011). unmarked: An R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software*, 43(10):1–23.

Flegal, J. M., Hughes, J., Vats, D., and Dai, N. (2020). *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, Denver, CO, Coventry, UK, and Minneapolis, MN. R package version 1.4-1.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007). Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis*, 51(7):3509–3528.

Frühwirth-Schnatter, S. and Frühwirth, R. (2010). Data augmentation and MCMC for binary and multinomial logit models. In Kneib, T. and Tutz, G., editors, *Statistical Modelling and Regression Structures*, pages 111–132. Physica-Verlag, Heidelberg.

Gardner, C. L., Lawler, J. P., Ver Hoef, J. M., Magoun, A. J., and Kellie, K. A. (2010). Coarse-scale distribution surveys and occurrence probability modeling for wolverine in interior Alaska. *Journal of Wildlife Management*, 74(8):1894–1903.

Gelfand, A. E., Schmidt, A. M., Wu, S., Silander, J. A., Latimer, A., and Rebelo, A. G. (2005). Modelling species diversity through species level hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):1–20.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, 4(1):11–15.

Gelling, N., Schofield, M. R., and Barker, R. J. (2019). R package rjmcmc: reversible jump mcmc using post-processing. *Australian and New Zealand Journal of Statistics*, 61(2):189–212.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian analysis*, 1(3):515–534.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Gershman, S., Hoffman, M., and Blei, D. (2012). Nonparametric variational inference. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, page 663–670.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian Statistics*, 4:641–649.

Ghahramani, Z. and Beal, M. J. (2000). Graphical models and variational methods. In *Advanced Mean Field Methods - Theory and Practice*. MIT Press.

Goijman, A. P., Conroy, M. J., Bernardos, J. N., and Zaccagnini, M. E. (2015). Multi-season regional analysis of multi-species occupancy: implications for bird conservation in agricultural lands in east-central argentina. *PloS one*, 10(6):e0130874.

Golub, G. H. and Welsch, J. H. (1969). Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106):221–230.

Gosselin, F. (2011). A New Calibrated Bayesian Internal Goodness-of-Fit Method: Sampled Posterior p-Values as Simple and General p-Values That Allow Double Use of the Data. *PloS one*, 6(3):e14770.

Grimmer, J. (2010). An introduction to Bayesian inference via variational approximations. *Political Analysis*, 19(1):32–47.

Guillera-Arroita, G., Lahoz-Monfort, J. J., MacKenzie, D. I., Wintle, B. A., and McCarthy, M. A. (2014). Ignoring Imperfect Detection in Biological Surveys Is Dangerous: A Response to 'Fitting and Interpreting Occupancy Models'. *PloS One*, 9(7):e99571.

Hanks, E. M., Hooten, M. B., and Baker, F. A. (2011). Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecological applications*, 21(4):1173–1188.

Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254.

Harebottle, D. M., Smith, N., Underhill, L. G., and Brooks, M. (2007). Southern African bird atlas project 2: Instruction manual. *Animal Demography Unit, University of Cape Town, Cape Town*.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC Press.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hensman, J., Rattray, M., and Lawrence, N. D. (2012). Fast variational inference in the conjugate exponential family. In Bartlett, P. L., Pereira, F. C. N., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 25, Cambridge, MA.

Hjort, N. L., Dahl, F. A., and Steinbakk, G. H. (2006). Post-processing posterior predictive p values. *Journal of the American Statistical Association*, 101(475):1157–1174.

Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334.

Hoeting, J. A., Leecaster, M., and Bowden, D. (2000). An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(1):102–114.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Statistical Science*, 14(4):382–417.

Hoffman, M. D. and Gelman, A. (2014). The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1):1593–1623.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis*, 1(1):145–168.

Hooten, M. B., Larsen, D. R., and Wikle, C. K. (2003). Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, 18(5):487–502.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):139–159.

Hugo, S. and Altwegg, R. (2017). The second southern african bird atlas project: Causes and consequences of geographical sampling bias. *Ecology and evolution*, 7(17):6839–6849.

Huntley, B., Collingham, Y. C., Green, R. E., Hilton, G. M., Rahbek, C., and Willis, S. G. (2006). Potential impacts of climatic change upon geographical distributions of birds. *Ibis*, 148(s1):8–28.

Hutchinson, R. A., Valente, J. J., Emerson, S. C., Betts, M. G., and Dietterich, T. G. (2015). Penalized likelihood methods improve parameter estimates in occupancy models. *Methods in Ecology and Evolution*, 6(8):949–959.

Jaakkola, T. S. and Jordan, M. I. (2000). Bayesian logistic regression: a variational approach. *Statistics and Computing*, 10(2):25–37.

Johnson, D. S., Conn, P. B., Hooten, M. B., Ray, J. C., and Pond, B. A. (2013). Spatial occupancy models for large data sets. *Ecology*, 94(4):801–808.

Joseph, Maxwell, B. (2012). A step-by-step guide to marginalizing over discrete parameters for ecologists using Stan. `https://mbjoseph.github.io/posts/2020-04-28-a-step-by-step-guide-to-marginalizing-over-discrete-parameters-for-ecologists-using-stan/`. [Accessed on 20 September 2021].

Kellner, K. (2014). jagsui: Run JAGS (specifically, libjags) from R; an alternative user interface for rjags. *R package version*, 1.

Kelsall, J. and Wakefield, J. (1999). Contribution to: "Bayesian models for spatially correlated disease and exposure data", by Best, N.G., Waller, L.A., Thomas, A., Conlon, E.M., and Arnold, R. In *Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting*, volume 151. Oxford University Press: Oxford.

Kéry, M., Gardner, B., and Monnerat, C. (2010). Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, 37(10):1851–1862.

Kinnaird, M. F. and O'Brien, T. G. (2012). Effects of private-land use, livestock management, and human tolerance on diversity, distribution, and abundance of large African mammals. *Conservation Biology*, 26(6):1026–1039.

Kuhnert, P. M., Martin, T. G., and Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in Bayesian ecological models. *Ecology Letters*, 13(7):900–914.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of mathematical statistics*, 22(1):79–86.

Latimer, A. M., Wu, S., Gelfand, A. E., and Silander, J. A. (2006). Building statistical models to analyze species distributions. *Ecological applications*, 16(1):33–50.

Lichstein, J. W., Simons, T. R., Shriner, S. A., and Franzreb, K. E. (2002). Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72(3):445–463.

Link, W. A. and Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, 3(1):112–115.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.

MacKay, D. J. (1995). Developments in probabilistic modelling with neural networks-ensemble learning. In *Neural Networks: Artificial Intelligence and Industrial Applications*, pages 191–198. Springer.

MacKenzie, D. I. (2006). *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Elsevier/Academic Press.

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., and Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84(8):2200–2207.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

MacKenzie, D. I., Nichols, J. D., Seamans, M. E., and Gutiérrez, R. (2009). Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, 90(3):823–835.

Martin, J., McIntyre, C. L., Hines, J. E., Nichols, J. D., Schmutz, J. A., and MacCluskie, M. C. (2009). Dynamic multistate site occupancy models to evaluate hypotheses relevant to conservation of golden eagles in denali national park, alaska. *Biological Conservation*, 142(11):2726–2731.

McClintock, B. T., Bailey, L. L., Pollock, K. H., and Simons, T. R. (2010). Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, 91(8):2446–2454.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

Mcfadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In Zarembka, P., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press, New York.

McGrory, C. A. and Titterington, D. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics and Data Analysis*, 51(11):5352–5367.

Meng, X.-L. et al. (1994). Posterior predictive p-values. *The Annals of Statistics*, 22(3):1142–1160.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Miller, D. A., Nichols, J. D., McClintock, B. T., Grant, E. H. C., Bailey, L. L., and Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, 92(7):1422–1428.

Mohankumar, N. M. and Hefley, T. J. (2021). Using machine learning to model nontraditional spatial dependence in occupancy data. *Ecology*, page e03563.

Monnahan, C. C., Thorson, J. T., and Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, 8(3):339–348.

Moreno, M. and Lele, S. R. (2010). Improved estimation of site occupancy using penalized likelihood. *Ecology*, 91(2):341–346.

Nathoo, F. S., Babul, A., Moiseev, A., Virji-Babul, N., and Beg, M. F. (2014). A variational Bayes spatiotemporal model for electromagnetic brain mapping. *Biometrics*, 70(1):132–143.

Nelder, J. and Wedderburn, R. (1972). Generalized linear model. *Journal of the Royal Statistical Society*, 135(3):370–384.

Ormerod, J. T. (2011). Skew-Normal Variational Approximations for Bayesian Inference. Technical report, School of Mathematics and Statistics, University of Sydney, Technical Report CRG- TR-93-1.

Ormerod, J. T. and Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, 62(2):140–153.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife monographs*, (62):3–135.

Péron, G. and Altwegg, R. (2015). The abundant centre syndrome and species distributions: insights from closely related species pairs in southern Africa. *Global Ecology and Biogeography*, 24(2):215–225.

Peterman, W. E., Crawford, J. A., and Kuhns, A. R. (2013). Using species distribution and occupancy modeling to guide survey efforts and assess species status. *Journal for Nature Conservation*, 21(2):114–121.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In Hornik, K., Leisch, F. and Zeileis, A., editor, *Proceedings of the 3rd international workshop on distributed statistical computing*. Technische Universit at Wien, Vienna, Austria.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Plummer, M., Best, N., Cowles, K., Vines, K., and Plummer, M. M. (2007). The coda package. *International Agency for Research on Cancer, France.[cited 30 December 2004.] Available from URL: http://www-fis. iarc. fr/coda*.

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., and McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5):397–406.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349.

Ponisio, L. C., de Valpine, P., Michaud, N., and Turek, D. (2020). One size does not fit all: Customizing mcmc methods for hierarchical models using nimble. *Ecology and evolution*, 10(5):2385–2416.

Qi, J., Chehbouni, A., Huerte, A., Kerr, Y., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2):119–126.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A. E. (1993). Bayesian model selection in structural equation models. *Sage Focus Editions*, 154:163–163.

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, 62(4):1197–1206.

Reynolds, D. (2015). Gaussian mixture models. *Encyclopedia of biometrics*, pages 827–832.

Robert, C. and Casella, G. (1999). *Monte Carlo statistical methods*. Springer texts in statistics. Springer.

Roberts, G. O. and Sahu, S. K. (1997). Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):291–317.

Roth, T., Strebel, N., and Amrhein, V. (2014). Estimating unbiased phenological trends by adapting site-occupancy models. *Ecology*, 95(8):2144–2154.

Royle, J. A. and Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Academic Press, San Diego, CA.

Royle, J. A. and Kery, M. (2007). A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, 88(7):1813–1823.

Royle, J. A. and Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4):835–841.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Sethodology)*, 71(2):319–392.

Ruiz-Gutiérrez, V., Zipkin, E. F., and Dhondt, A. A. (2010). Occupancy dynamics in a tropical bird community: unexpectedly high forest use by birds classified as non-forest species. *Journal of Applied Ecology*, 47(3):621–630.

Sanderson, C. and Curtin, R. (2016). Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software*, 1(2):26.

Sauer, J. R., Blank, P. J., Zipkin, E. F., Fallon, J. E., and Fallon, F. W. (2013). Using multi-species occupancy models in structured decision making on managed lands. *The Journal of Wildlife Management*, 77(1):117–127.

Seber, G. A. (2008). *A matrix handbook for statisticians*, volume 15. John Wiley and Sons.

Smith, B. J. (2007). boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, 21(11):1–37.

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2007). OpenBUGS user manual, version 3.0. 2. *MRC Biostatistics Unit, Cambridge*.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A. (1998). Bayesian Deviance, the Effective Number of Parameters, and the Comparison of Arbitrarily

Complex Models. Technical Report 98-009, Division of Biostatistics, University of Minnesota. Cambridge, UK: Medical Research Council Biostatistics University.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stan Development Team (2020a). RStan: the R interface to Stan. R package version 2.21.2.

Stan Development Team (2020b). Stan Modeling Language Users Guide and Reference Manual. R package version 2.21.0.

Su, Y. and Yajima, M. (2012). R2jags: A Package for Running jags from R. [Accessed on 04 January 2015].

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Thomas, A., O'Hara, B., Ligges, U., and Sturtz, S. (2006). Making BUGS open. *R news*, 6(1):12–17.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M., Oksanen, J., and Ovaskainen, O. (2020). Joint species distribution modelling with the R-package Hmsc. *Methods in ecology and evolution*, 11(3):442–447.

Tobler, M. W., Zúñiga Hartley, A., Carrillo-Percastegui, S. E., and Powell, G. V. (2015). Spatiotemporal hierarchical modelling of species and occupancy using camera trap data. *Journal of Applied Ecology*, 52(2):413–421.

Turek, D., de Valpine, P., Paciorek, C. J., and Anderson-Bergman, C. (2017). Automated parameter blocking for efficient markov chain monte carlo sampling. *Bayesian Analysis*, 12(2):465–490.

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5):1413–1432.

Vieilledent, G., Merow, C., Guélat, J., Latimer, A., Kéry, M., Gelfand, A., Wilson, A., Mortier, F., and Silander Jr, J. (2014). hSDM: hierarchical Bayesian species distribution models. R package version 1.4.

Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*. Wiley, New York.

Wang, B. and Titterington, D. M. (2005). *Inadequacy of interval estimates corresponding to variational Bayesian approximations*. Proceedings of the 10th International Workshop on Artificial Intelligence.

Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:3571–3594.

Welsh, A. H., Lindenmayer, D. B., and Donnelly, C. F. (2013). Fitting and interpreting occupancy models. *PLoS One*, 8(1):e52015.

Wilkinson, D. (2005). Handbook of parallel computing and statistics, chapter parallel bayesian computation.

Williams, C. K. and Rasmussen, C. E. (2006). *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA.

Windle, J., Polson, N., and Scott, J. (2013). BayesLogit: Bayesian logistic regression. *URL http://cran. r-project. org/web/packages/BayesLogit/index. html. R package version 0.2-4*.

Wright, W. J., Irvine, K. M., and Higgs, M. D. (2019). Identifying occupancy model inadequacies: can residuals separately assess detection and presence? *Ecology*, 100(6):e02703.

Yackulic, C. B., Dodrill, M., Dzul, M., Sanderlin, J. S., and Reid, J. A. (2020). A need for speed in bayesian population models: a practical guide to marginalizing and recovering discrete latent states. *Ecological Applications*, 30(5):e02112.

You, C., Ormerod, J. T., and Müller, S. (2014). On variational Bayes estimation and variational information criteria for linear regression models. *Australian and New Zealand Journal of Statistics*, 56(1):73–87.

Zhang, J. L. (2014). Comparative investigation of three Bayesian p values. *Computational Statistics and Data Analysis*, 79:277–291.

Zipkin, E. F., DeWan, A., and Andrew Royle, J. (2009). Impacts of forest fragmentation on species richness: a hierarchical approach to community modelling. *Journal of Applied Ecology*, 46(4):815–822.

Zipkin, E. F., Royle, J. A., Dawson, D. K., and Bates, S. (2010). Multi-species occurrence models to evaluate the effects of conservation and management actions. *Biological Conservation*, 143(2):479–484.

Zobay, O. (2014). Variational Bayesian inference with Gaussian-mixture approximations. *Electronic Journal of Statistics*, 8(1):355–389.