Is anybody home? Modelling frog occupancy at the La Trobe Wildlife Sanctuary.

Kevin Newman

Supervised by Natalie Karavarsamis and Hien Nguyen
La Trobe University

Is anybody home? Modelling frog occupancy at the La Trobe Wildlife Sanctuary.

K. Newman[*] and N. Karavarsamis

School of Mathematics and Statistics

La Trobe University, Victoria 3086, Australia.

ABSTRACT: Occupancy models are a sound statistical process for determining dispersion and abundance of species. These models allow us to estimate the probability of a site being occupied and the probability of detection at an occupied site for a species of interest. As a cheaper and less labour intensive alternative to capture-recapture methods these models can assist in improving conservation efforts to protect the future of populations. The full likelihood and two-stage approaches are analysed and applied to detection data collected as part of the Melbourne Water Frog Census at the La Trobe University Wildlife Sanctuary from 2014 - 2018. Both methods were applied first in the homogeneous case (without covariates) before being extended to include covariate information in the heterogeneous case. These approaches did not always produce meaningful results due to the low detection probabilities of some species. Simulations were run to compare the two methods which were found to be consistent for occupancy probability, with some variation in detection probabilities. As detection is indicative of human behaviour it is of less conservation importance than occupancy. A multiseason approach which takes into account the probability of colonisation and extinction showed that four out of the five species of frogs present in the wildlife sanctuary had trends of declining occupancy probabilities.

KEY WORDS: Imperfect detection; Occupancy models; Orthogonal parameterisation; Partial likelihood.

1 Introduction

When conducting ecological experiments and studies in the field the likelihood of detecting a species in their natural environment often has a probability < 1. For many cases, non-detection at a site does not infer that the species is not present but may be due to other covariates. MacKenzie et al. (2002) proposed a model and likelihood-based method for estimating occupancy rates for species with detection rates of < 1 which has been developed further into a two-stage approach by Karavarsamis and Huggins (2018a). Unlike many current methods in ecology occupancy models can allow investigators to define their own sample units, time frame and criteria for detection and do not rely on capture-recapture. In comparison, methods that work on presence and absence data are less costly and more time efficient.

Many conservation efforts, such as those of avian species in the South Atlantic Coastal Plain of the United States (Eastern Wood-Pewee [*Contropus virens*], Brown-headed Nuthatch [*Sitta pusilla*] & Red-headed Woodpecker [*Malenerpes erythrocephalu*]), already consider the scarce amount of data we have on species-habitat relationships (Iglecia et al., 2012). Improved methods to obtain accurate information on the dispersion of species ensure that we would be making appropriate decisions regarding conservation efforts to protect the future of these populations. Occupancy models can contribute to the broader understanding of ecosystems and are a sound process for statistically determining abundance and dispersion, and so are critical for the future of species and ecosystems (Hoffman et al., 2010). Studying dispersion of a species is cheaper and less labour intensive than capture-recapture methods, as it only requires presence/absence data to be generated in the survey.

Initially, this project will analyse the basic occupancy model as described by MacKenzie et al. (2002) and its associated R package, unmarked (R Development Core Team, 2018; Fiske and Chandler, 2015). Using simulations, we will compare the approaches of MacKenzie et al. (2002); Karavarsamis and Huggins (2018a). These two approaches will be applied to data obtained of Frog Species from the La Trobe Wildlife Sanctuary Frog Census monitoring program and will also be expanded to include data on covariates and the impact that this has on resulting output. This data lends itself well to occupancy modelling as it relies on detection of frog calls on a number of survey periods at a set number of sites.

In Section 2 we give our notation and in Section 3 the full likelihood of MacKenzie et al. (2002) and two-stage approach of Karavarsamis and Huggins (2018a) to homogeneous models. In Section 4 we extend the model to the heterogeneous case and include information on covariates. We describe our data in Section 5 and in Section 6 we apply the estimation methods. In Section 7 we conduct some

simulations to examine the efficiency of the resulting occupancy estimators. Some technical results are given in the Appendices.

2 Notation

We denote S sites labelled $s = 1, \ldots, S$ and each site surveyed on τ occasions. It is supposed that occupancy is constant for each survey. We denote the probability of a site being occupied by ψ and p as the probability of detection, the species is recorded at a site on a survey occasion given that the species is present at the site. The probability of at least one detection, given that a site is occupied can be calculated by $\theta = 1 - (1 - p)^{\tau}$. We let y_s denote the number of surveys in which a species was detected at a particular site s and $y = \sum_{s=1}^{S} y_s$ be the total number of detections across all sites, f_0 is defined as sites which no detections occurred and the number of sites which detections occurred are $O = S - f_0$. Survey occasion is denoted by j and p_j is the probabiity of detection at time j. We reorder the S sites $s = 1, \ldots, O, O+1, \ldots S$, so that $1, \ldots, O$ denote sites where at least one detection occurred and $O+1, \ldots S$ denotes the sites at which no detections occurred. For the multiseason model we define colonisation by γ, extinction by ϵ and survey period (year) by T.

3 Homogeneous models

MacKenzie et al. (2002) describes the method of estimating site occupancy of a species as a consideration for large scale monitoring programs or to understand metapopulation dynamics. The method was developed to allow for estimation of site occupancy when detection probabilities of the species is < 1. This does not necessarily mean that a species is not present, just that they were not detected on that particular occasion.

The MacKenzie et al. (2002) method is similar to a closed-population model of capture-recapture but includes the addition of a new parameter (ψ) which is representative of species occupancy. Detection of a species at a site is assumed independent of all other sites and resulting data is recorded as a vector of 0's (non-detection) and 1's (detection).

Using a series of probabilistic arguments and assuming that sites are independent the following equation gives the model likelihood for the observed set of data which can then be maximised to obtain the maximum likelihood estimates for the parameters. The full likelihood (MacKenzie et al., 2002) is

$$L(\psi, p) = \left[\psi^O \prod_{j=1}^{\tau} p_j^{y_s} (1 - p_j)^{O - y_s} \right] \times \left[\psi \prod_{j=1}^{\tau} (1 - p_j) + (1 - \psi) \right]^{S - O} \tag{1}$$

Equation 1 may be maximised numerically and the maximum likelihood estimates $\hat{\psi}$ and $\hat{p}$ found simultaneously with the R package `unmarked` (Fiske and Chandler, 2015)

The model by MacKenzie et al. (2002) assumes that: occupancy state of survey units is constant across a survey period, probability of occupancy is equal across all survey sites, the probability of detection, given occupancy, is equal across sites for all survey periods, detections and detection histories are independent and there are no false absences.

The two-stage approach that has been developed by Karavarsamis and Huggins (2018a) simplifies estimation by exploiting that data collected provides more information on detection probabilities than it does for occupancy probabilities. This approach ignores the information up to and including the first detection at each site where b_s is the number of occasions remaining after the first detection at site s and $b = \sum_{s=1}^{S} b_s$. The number of re-detections at site s is $y_s - 1$ so that the total number of re-detections is $y - O$.

Using the partial likelihood to estimate p yields that the partial likelihood estimator of p is $\tilde{p} = (y - O)/b$ which has a usual Binomial variance $\text{Var}(\tilde{p}) = \tilde{p}(1 - \tilde{p})/b$. We then estimate ψ by first estimating η and then back transform to yield $\widetilde{\psi} = \hat{\eta}/\widetilde{\theta} = (S - f_0)/S\widetilde{\theta}$, where $\widetilde{\theta} = 1 - (1 - \tilde{p})^\tau$. The equation for the variance of $\hat{\psi}$ can then be given as

$$\text{Var}\left(\widetilde{\psi}\right) \approx \left(\frac{\psi(1 - \psi\theta)}{S\theta} + \psi^2\right)\frac{\tau^2(1 - p)^{2(\tau-1)}}{\theta^2}\frac{p(1 - p)}{b} + \frac{\psi(1 - \psi\theta)}{S\theta}. \tag{2}$$

We can estimate p using the R package `VGAM` then use these estimates (now fixed numbers) to estimate ψ in the second stage via an Iterative Weighted Least Squares (IWLS) method until there is convergence to an estimate.

4   Heterogeneous models

Often models are not as simple as Section 3 and we need to generalise to allow for heterogeneity in detection across sites and surveys. Due to their variable nature, factors that influence detectability may be difficult to identify or control for (Karavarsamis, 2015). Variation in detection may be influenced by temporal covariates, such as changes in temperature or rainfall over time, or site specific covariates such as vegetation class or region within the study location.

The models of Section 3 assume that ψ and p are constant for site and survey occasions, but of more interest and relevance to practical applications is allowing occupancy and detection to vary in accordance with some covariates. Thus the models developed by MacKenzie et al. (2002) and

Karavarsamis and Huggins (2018a) can be extended to include covariate information and best fitting models can be found to estimate occupancy and detection probabilities.

For the heterogeneous case we modify the equation for the full likelihood so that the contribution of site s is

$$L_s(\psi_s, p_s) = (1 - \psi_s + \psi_s(1 - p_s)^\tau)^{Z_s} \left\{ \binom{\tau}{y_s} \psi_s p_s^{y_s} (1 - p)^{\tau - y_s} \right\}^{1 - z_s}$$

$$\propto (1 - \psi_s \theta_s)^{z_s} \psi_s^{1 - z_s} \{p_s(1 - p_s)^{(a_s - 1)}\}^{(1 - z_s)} \{p_s^{(y_s - 1)}(1 - p)^{b_s - y_s + 1}\}^{(1 - z_s)}.$$

$$= L_{1s}(\psi_s, p_s) L_{2s}(p_s) \tag{3}$$

where

$$L_{2s}(p_s) = \{p_s^{(y_s - 1)}(1 - p)^{b_s - y_s + 1}\}^{(1 - z_s)} \tag{4}$$

and

$$L_{1s}(\psi_s, p_s) = (1 - \psi_s + \psi_s(1 - p_s)^\tau)^{Z_s} \psi_s^{1 - z_s} \{p_s(1 - p_s)^{(a_s - 1)}\}^{(1 - z_s)}$$

$$= (1 - \psi_s \theta_s)^{z_s} \psi_s^{1 - z_s} \{p_s(1 - p_s)^{(a_s - 1)}\}^{(1 - z_s)} \tag{5}$$

where $Z_s = \mathrm{I}(y_s = 0)$, the indicator function for no detections at site s and a_s is the first time in which a detection was made.

It is common for covariates for detection or occupancy to be related to each site so we suppose that $\psi_s = h(x_s^T \boldsymbol{\alpha})$ where x_s is the vector of coefficients that are relevant to site s and $\boldsymbol{\alpha} \in \mathbb{R}^p$ is the vector of coefficients. We then let p_{sj} denote the probability that a detection is made at site s at survey j if occupied. We then take $p_{sj} = h(u_{sj}^T \boldsymbol{\beta})$ where u_{sj} is the covariate vector for site s on occasion j.

Covariates can either be included as time dependent or independent and their contribution to detection and occupancy probabilities is explored in Karavarsamis and Huggins (2018b) for the two-stage approach. In single season models occupancy is considered to be constant across all survey periods, thereby once a site is occupied it remains occupied, as such covariates which influence occupancy are also time independent whereas covariates which influence detection can be either (Karavarsamis and Huggins, 2018b).

5   Data

Detections were recorded at The La Trobe Wildlife Sanctuary, as part of the Melbourne Water Frog Census (Catus-Wood, 2017). Five frog species were studied, the Common Froglet [*Crinia signifera*],

Eastern Banjo Frog [*Limnodynastes dumerilii*], Spotted Marsh Frog [*Limnodynastes tasmanienses*], Southern Brown Tree Frog [*Litoria ewingii*] and Peron's Tree Frog [*Litoria peronii*].

The application utilises data collected at $S = 16$ sites which are shown in figure 1 as red stars with three or four surveys each year, $\tau = (3, 4, 4, 3, 4)$ over five years between 2014 and 2018. The data consists of a matrix of 1's (detections) and 0's (non-detections) alongside site names and covariate information. The frog census forms part of the La Trobe University Wildlife Sanctuary's citizen science program and occurs up to four times per year generally April or May, August or September, October, and November or December.
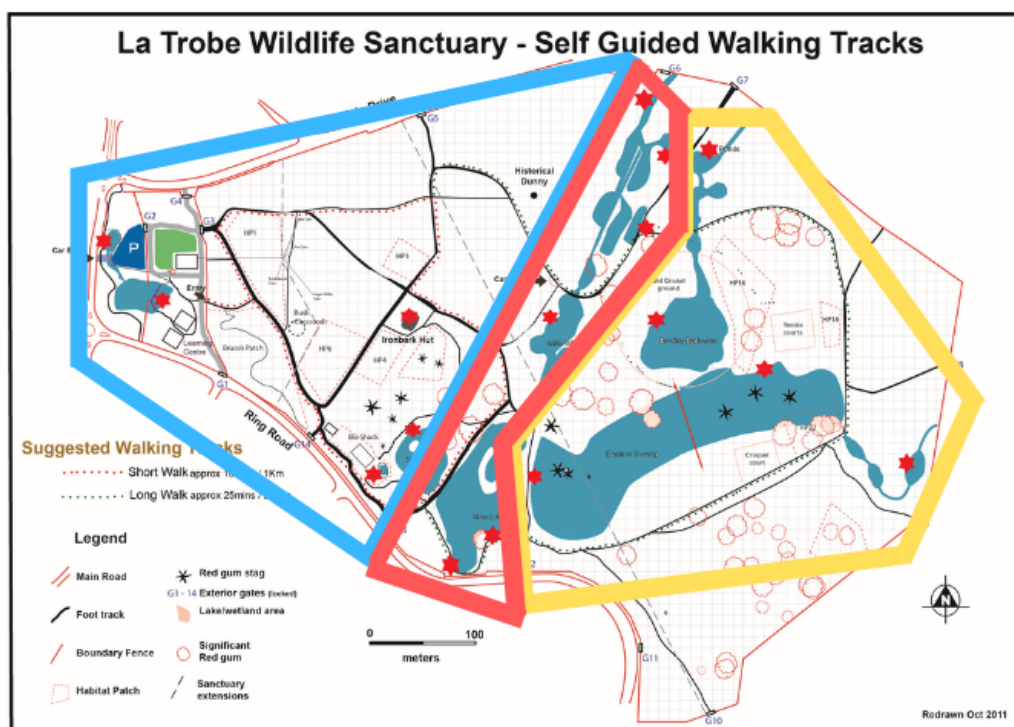

Figure 1: Map of the La Trobe University Wildlife Sanctuary. Survey sites are marked by red stars. We divided the sites into three regions based on location within the sanctuary, east (yellow), central (red) and west (blue).

6 Single species and single season model

Initially, in Appendix A we conducted a small modelling exercise to examine the assumption of closed population and then calculated the estimates for occupancy, detection, colonisation and extinction from the most parsimonious models using the methods of MacKenzie et al. (2003). Overall, probability of colonisation and extinction are low, and in effect practically negligible on the assumption of closed populations.

6.1 Homogeneous

Here we consider the full likelihood approach and two-stage approach for homogeneous models. To compare the full and two-stage methods for a single species and single season we utilise data for each species from the 2014 - 2018 survey period. Efficiencies were calculated as standard ratio of variances to compare the two methods.

Results for *C. signifera* are reported in Table 1. Estimates were above 0.75 for both detection and occupancy probability. Occupancy from the full and two stage approach give estimates that are equal to two significant figures. For some years detection estimates vary between full and two stage methods however does not affect accuracy of occupancy estimation and is of little concern for species planning.

Table 1: Single season approach applied to individual survey years for *C. signifera*. Estimates for occupancy $\hat{\psi}$ and detection $\hat{p}$ are calculated with the two-stage approach and `occu` and their associated standard errors.

		Full		Two-stage		
		EST	SE	EST	SE	Efficiency
2014	$\hat{\psi}$	0.885	0.084	0.887	0.085	0.977
	$\hat{p}$	0.777	0.068	0.760	0.085	0.640
2015	$\hat{\psi}$	0.814	0.098	0.813	0.098	1.000
	$\hat{p}$	0.807	0.055	0.829	0.064	0.739
2016	$\hat{\psi}$	0.938	0.061	0.938	0.061	1.000
	$\hat{p}$	0.850	0.046	0.900	0.047	0.958
2017	$\hat{\psi}$	0.750	0.108	0.750	0.108	1.000
	$\hat{p}$	0.916	0.047	0.913	0.059	0.635
2018	$\hat{\psi}$	0.814	0.098	0.813	0.098	1.000
	$\hat{p}$	0.787	0.058	0.875	0.058	1.000

Table 2 reports the estimates for detection and occupancy for *L. ewingii*. Detection estimates are relatively low (between 0.30 and 0.65) and the estimates for occupancy range between 0.60 and 1.01. Efficiencies are above 65% for both occupancy and detection. Standard errors are large when considered in context of the estimates and this may impact the accuracy of estimation.

Table 2: Single season approach applied to individual survey years for *L. ewingii*. Estimates for occupancy $\hat{\psi}$ and detection $\hat{p}$ are calculated with the two-stage approach and `occu` and their associated standard errors.

		Full		Two-stage		
		EST	SE	EST	SE	Efficiency
2014	$\hat{\psi}$	0.691	0.144	0.678	0.151	0.909
	$\hat{p}$	0.543	0.110	0.571	0.132	0.694
2015	$\hat{\psi}$	0.991	0.134	1.012	0.163	0.676
	$\hat{p}$	0.347	0.076	0.333	0.091	0.698
2016	$\hat{\psi}$	0.932	0.178	0.954	0.182	0.957
	$\hat{p}$	0.335	0.086	0.320	0.093	0.855
2017	$\hat{\psi}$	0.789	0.118	0.857	0.146	0.653
	$\hat{p}$	0.634	0.092	0.500	0.102	0.814
2018	$\hat{\psi}$	0.662	0.163	0.639	0.172	0.898
	$\hat{p}$	0.378	0.099	0.412	0.119	0.692

Table 3 presents the estimates for occupancy and detection for *L. dumerilii*. We see that both methods have resulted in occupancy probabilities that are greater than one and the two-stage approach in a detection probability estimate equal to 1 for the 2018 survey period as well as non-convergence in the full likelihood method. *L. dumerilii* are seasonal breeders, with their main calling period in the spring and this will influence detection. The data used for this application was collected four times a year, with only the final two survey points during the appropriate season for detection. Particularly for the two-stage approach, as this ignores the first survey in which detection occurs, we lose a lot of information and a different survey schedule would need to be implemented to get an accurate picture.

In table 4 we present the detection and occupancy probabilities for *L. tasmaniensis*. The results for this application vary greatly between the full and two-stage approaches, with the full method hitting the boundary of 1 for occupancy in 2014, 2015 and 2018.

For *L. peronii* the full likelihood method produced estimates for 2016 ($\hat{\psi} = 0.999$ with SE 0.093

Table 3: Single season approach applied to individual survey years for *L. dumerilii*. Estimates for occupancy $\hat{\psi}$ and detection $\hat{p}$ are calculated with the two-stage approach and `occu` and their associated standard errors. DNC indicates that the model did not converge to an estimate.

		Full		Two-stage		
		EST	SE	EST	SE	Efficiency
2014	$\hat{\psi}$	0.941	0.243	0.753	0.161	2.278
	$\hat{p}$	0.354	0.114	0.556	0.166	0.472
2015	$\hat{\psi}$	1	0.008	1.817	0.530	0.0002
	$\hat{p}$	0.172	0.047	0.100	0.095	0.245
2016	$\hat{\psi}$	0.984	0.289	0.647	0.139	4.323
	$\hat{p}$	0.222	0.083	0.571	0.187	0.197
2017	$\hat{\psi}$	0.908	0.349	0.744	0.263	1.761
	$\hat{p}$	0.275	0.124	0.375	0.171	0.526
2018	$\hat{\psi}$	DNC		0.375	0.121	-
	$\hat{p}$	DNC		1	0	-

and $\hat{p} = 0.031$ with standard error 0.022) and 2017 ($\hat{\psi} = 0.483$ with SE = 0.371 and $\hat{p}$ 0.215 and SE = 0.178). The species has low detection probability within the study area, and neither method can provide reliable estimates as reflected in inflated standard errors for the full likelihood, which is known to underestimate variances. The two-stage approach did not converge.

Table 4: Single season approach applied to individual survey years for *L. tasmaniensis*. Estimates for occupancy $\hat{\psi}$ and detection $\hat{p}$ are calculated with the two-stage approach and `occu` and their associated standard errors.

		Full		Two-stage		
		EST	SE	EST	SE	Efficiency
2014	$\hat{\psi}$	1	0.007	0.891	0.146	0.002
	$\hat{p}$	0.375	0.070	0.556	0.166	0.178
2015	$\hat{\psi}$	0.880	0.248	0.859	0.233	1.133
	$\hat{p}$	0.266	0.092	0.278	0.106	0.753
2016	$\hat{\psi}$	1	0.008	0.969	0.069	0.013
	$\hat{p}$	0.469	0.063	0.577	0.097	0.422
2017	$\hat{\psi}$	0.606	0.139	0.567	0.126	1.217
	$\hat{p}$	0.584	0.111	0.800	0.126	0.776
2018	$\hat{\psi}$	1	0.016	0.504	0.218	0.005
	$\hat{p}$	0.141	0.044	0.500	0.354	0.015

6.2 Heterogeneous

We considered two site covariates; distance to closest walking track (Dist) and region defined as geographic location of a group of sites within the sanctuary (Reg). Model selection was run initially in `occu` for each individual survey year and species and then the most parsimonious model based on AIC values with the two-stage approach.

In the majority of instances the most parsimonious models were the original homogeneous models. In a few cases where region or distance were included non-convergence and inflated standard errors was evident, mostly on species with low detection probabilities. These are evident in Tables 17, 18, 19 and 20 of Appendix B. We can see that for some species detection is influenced by region or distance to the closest walking track but these methods are unreliable for our data. Like in the homogeneous case the models for *L. peronii* did not converge to estimates for both methods due to low detectability and these results were excluded.

7 Simulations

We ran a simulated experiment in order to compare the full likelihood and two-stage methods. We began by conducting 5000 simulations with $S = 100$, $\psi = 0.6$, and $p = 0.6$. In Table 5 we report the medians of the estimates, the associated standard errors and the median absolute deviation (MAD) for $\tau = 5$. The efficiency of the two-stage approach was calculated as a standard ratio of variances. Efficiency of the two-stage method for ψ is well above 90%.

We ran further simulations similar to our application by using $\tau = 18$ and $S = 16$ and these results are reported in Table 6. In this case the efficiency of the two-stage approach is above 90% for both p and ψ. Although efficiency of detection estimation is only 70% the accuracy of this probability has less of an impact on conservation management and species planning as detection is indicative of human behaviour rather than the behaviour of species and generally does not impact estimation of occupancy.

Table 5: Comparisons of the two-stage and the full likelihood estimates in the homogeneous case for a large number of sites ($S = 100$), small number of occasions ($\tau = 5$) and large values of p. We give the true values of the parameters, the median of the estimated values, the standard deviation and the median absolute deviation (MAD) of the estimates.

	Two-stage		Full	
	p	ψ	p	ψ
True Value	0.6000	0.6000	0.6000	0.6000
Median estimate	0.6000	0.5989	0.6004	0.5980
Standard deviation	0.0353	0.0501	0.0298	0.0500
MAD	0.0353	0.0511	0.0299	0.0501
Efficiency	0.7150	0.9982		

Table 6: Comparisons of the two-stage and the full likelihood estimates in the homogeneous case for a small number of sites ($S = 16$), large number of occasions ($\tau = 18$) and large values of p. We give the true values of the parameters, the median of the estimated values, the standard deviation and the median absolute deviation (MAD) of the estimates.

	Two-stage		Full	
	p	ψ	p	ψ
True Value	0.6000	0.6000	0.6000	0.6000
Median estimate	0.6000	0.6250	0.5987	0.6249
Standard deviation	0.0391	0.1219	0.0374	0.1219
MAD	0.0382	0.0927	0.0366	0.0928
Efficiency	0.9145	0.9999		

8   Conclusions

Species with low detectability cause issues when both `occu` and two stage are implemented in the homogeneous case. Including covariate information into heterogeneous models did not improve accuracy of estimation, but this is likely due to the covariate information present in the data. In order to compare the full and two-stage methods both methods were initially tested on the most parsimonious models from the model selection process based on AIC values. We discovered that many of the models did not converge, or allow the methods to run, which indicated issues with the covariate information that had been collected. As such, models which included the covariates of Temp and Season were excluded. Whilst covariate information was recorded at the time of survey, temperature, which could vary between sites was recorded as an overall temperature instead of individual site temperatures. Season is not going to change across sites either, so this resulted in the sites having the same values for these variables resulting in errors in estimation using both the MacKenzie et al. (2002) and Karavarsamis and Huggins (2018a) methods. As discussed in both section 3 and 4 estimating the probability of occupancy is reliant on the detection probabilities being estimated first in the two-stage approach, or simultaneously in the full likelihood approach.

Our estimates of detection can vary slightly between the two methods, but this has little impact on conservation as detection is indicative of human behaviour, rather than the behaviour of species and generally does not affect accuracy of occupancy estimation. We are more interested in the accuracy of our occupancy probabilities as this leads to a greater understanding of how a species is interacting with the environment and can lead to the development of habitat suitability maps based on the probability that a species occupies an area in relation to covariates such as elevation or vegetation cover.

Future work within this area could involve tailoring the surveys to each species specifically in order to improve accuracy in estimation. As many of them are more likely to call in October or November, completing more surveys in this period would allow for more accurate estimation of occupancy and detection probabilities. Site and time varying covariates would also need to be recorded accurately to ensure that they were able to provide more information into the models. The current surveying method is more suited to a multiseason approach, and therefore needs to be considered in the context of the four parameters which includes colonisation and extinction (MacKenzie et al., 2003). These models (Appendix A) show a decline in the occupancy probability of *L. dumerilii, L. peronii, L. ewingii* and *L. tasmaniensis*. Also, that the probability of colonisation and extinction are low for most species, so it is likely that this is a closed population and there is little to no movement between sites. Whilst our models suggest a decline in occupancy probability for species, we can not provide an explanation

as to why this is based on our covariate information but this does indicate that future work in this area could assist in conservation planning.

Acknowledgements

References

Catus-Wood, M. (2017). Hop to it - be a part of melbourne water's frog census program. *Interaction*, 45(2):17–19.

Fiske, I. J. and Chandler, R. B. (2015). *Overview of unmarked: an R package for the analysis of data from unmarked animals*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Hoffman, J., Aguilar-Amuchastegui, N., and Tyre, A. (2010). Use of simulated data from a oricess-based habitat model to evaliate methods for predicting species occurrence. *Ecography*, 33:656–666.

Iglecia, M. N., Collazo, J. A., and McKerrow, A. J. (2012). Use of occupancy models to evaluate expert knowledge-based species-habitat relationships. *Avian Conservation and Ecology*, 7(2):5.

Karavarsamis, N. (2015). *Methods for estimating occupancy*. University of Melbourne, Victoria, Australia.

Karavarsamis, N. and Huggins, R. (2018a). A two-stage approach to the analysis of occupancy data I. The Homogeneous case (analysis of occupancy data). Arxiv ID: 1804.08694.

Karavarsamis, N. and Huggins, R. (2018b). A two-stage approach to the analysis of occupancy data II. The Heterogeneous case (in press). *Computational Statistics and Data Analysis*. DOI: `https://doi.org/10.1016/j.csda.2018.09.009`, Arxiv ID: 1803.11354.

MacKenzie, D., Nichols, J., Hines, J., Knutson, M., and Franklin, A. (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., Royle, J., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

A   Single species and multi season model - Colonization and extinction

Here we examine whether the assumption of a closed population is valid. This application also uses the same data as described in Section 5, however three survey occasions were removed (April in 2015, 2016 and 2019) in order to standardise the data to $T = 5$ and $j = 3$ so that the `colext` command could be utilised as this requires the same number of survey occasions in each year. Support for alternative models was assessed using AIC, results can be found in Tables 7, 9, 11, 13 and 15. Overall, probability of colonisation and extinction are low, and in effect practically negligible on the assumption of closed populations (Tables 8, 10, 12, 14 and 16). Aside from *C. signifera* we note that these results show declines in occupancy rates for the remaining species.

Table 7: Multiseason models of occupancy (ψ) for *C. signifera.* We present the AIC and distance from the most parsimonious model (ΔAIC).

ψ	γ	ϵ	p	AIC	ΔAIC
Int	Int	Int	Reg	221.05	0
Reg	Int	Int	Reg	223.44	2.39
Int	Int	Int	Int	224.91	3.86
Dist	Int	Int	Int	225.70	4.65
Reg	Int	Int	Int	227.30	6.25
Int	Int	Int	Year	227.80	6.75
Int	Year	Year	Int	231.51	10.46
Int	Year	Year	Year	234.42	13.37
Dist	Int	Int	Temp	255.70	34.65
Reg	Int	Int	Temp	257.30	36.25
Temp	Int	Int	Year	264.10	43.05

Table 8: Multiseason approach applied to the most parsimonious model from Table 7. Here we present the coefficient estimates for occupancy ($\hat{\psi}$), colonisation (γ), extinction (ϵ) and detection ($\hat{p}$) calculated with `colext` and their associated standard errors

	EST	SE		EST	SE		EST	SE
ψ_{2014}	0.877	0.055	$p_{central}$	0.904	0.033	ϵ	0.074	0.041
ψ_{2015}	0.844	0.077	p_{east}	0.893	0.038	γ	0.303	0.163
ψ_{2016}	0.875	0.077	p_{west}	0.701	0.078			
ψ_{2017}	0.785	0.091						
ψ_{2018}	0.815	0.063						

Table 9: Multiseason models of occupancy (ψ) for *L. dumerilii*. We present the AIC and distance from the most parsimonious model (ΔAIC).

ψ	γ	ϵ	p	AIC	ΔAIC
Dist	Int	Int	Temp	242.67	0.00
Reg	Int	Int	Temp	243.10	0.43
Int	Int	Int	Int	252.95	10.28
Dist	Int	Int	Int	254.95	12.28
Reg	Int	Int	Int	256.10	13.43
Int	Int	Int	Reg	256.16	13.49
Int	Int	Int	Year	256.82	14.15
Reg	Int	Int	Reg	259.32	16.65
Int	Year	Year	Int	262.65	19.99
Int	Year	Year	Year	267.34	24.67
Temp	Int	Int	Year	297.43	54.76

Table 10: Multiseason approach applied to the most parsimonious model from Table 9. Here we present the coefficient estimates for occupancy ($\hat{\psi}$), colonisation (γ), extinction (ϵ) and detection ($\hat{p}$) calculated with `colext` and their associated standard errors. The estimate for detection is calculated when temperature is fixed at its mean value of 14.6 °C and occupancy is for sites close to path.

	EST	SE		EST	SE		EST	SE
ψ_{2014}	0.755	0.076	p	0.073	0.039	ϵ	0.037	0.051
ψ_{2015}	0.751	0.077				γ	0.001	0.038
ψ_{2016}	0.725	0.085						
ψ_{2017}	0.678	0.106						
ψ_{2018}	0.650	0.130						

Table 11: Multiseason models of occupancy (ψ) for *L. peronii*. We present the AIC and distance from the most parsimonious model (ΔAIC).

ψ	γ	ϵ	p	AIC	ΔAIC
Int	Int	Int	Int	128.63	0.00
Int	Year	Year	Int	129.76	1.13
Dist	Int	Int	Int	130.65	2.01
Int	Int	Int	Year	131.69	3.06
Int	Int	Int	Reg	132.43	3.80
Reg	Int	Int	Int	132.62	3.99
Int	Year	Year	Year	135.30	6.66
Reg	Int	Int	Reg	135.38	6.74
Dist	Int	Int	Temp	160.67	32.03
Temp	Int	Int	Year	161.68	33.05
Reg	Int	Int	Temp	162.67	34.03

Table 12: Multiseason approach applied to the most parsimonious model from Table 11. Here we present the coefficient estimates for occupancy (ψ), colonisation (γ), extinction (ϵ) and detection (p) calculated with `colext` and their associated standard errors.

	EST	SE		EST	SE		EST	SE
ψ_{2014}	0.975	0.165	p	0.102	0.036	ϵ	0.166	0.123
ψ_{2015}	0.796	0.222				γ	0.002	0.052
ψ_{2016}	0.668	0.200						
ψ_{2017}	0.559	0.145						
ψ_{2018}	0.479	0.149						

Table 13: Multiseason models of occupancy (ψ) for *L. ewingii*. We present the AIC and distance from the most parsimonious model (ΔAIC).

ψ	γ	ϵ	p	AIC	ΔAIC
Int	Int	Int	Reg	305.04	0.00
Reg	Int	Int	Reg	305.81	0.77
Int	Int	Int	Int	307.39	2.35
Int	Year	Year	Int	308.43	3.39
Dist	Int	Int	Int	308.89	3.84
Reg	Int	Int	Int	309.16	4.12
Int	Int	Int	Year	310.39	5.35
Int	Year	Year	Year	314.42	9.38
Temp	Int	Int	Year	341.28	36.24
Dist	Int	Int	Temp	353.37	48.33
Reg	Int	Int	Temp	353.51	48.47

Table 14: Multiseason approach applied to the most parsimonious model from Table 13. Here we present the coefficient estimates for occupancy ($\hat{\psi}$), colonisation (γ), extinction (ϵ) and detection ($\hat{p}$) calculated with `colext` and their associated standard errors.

	EST	SE		EST	SE		EST	SE
ψ_{2014}	0.946	0.168	$p_{central}$	0.386	0.053	ϵ	0.061	0.039
ψ_{2015}	0.825	0.095	p_{east}	0.600	0.065	γ	0.0004	0.015
ψ_{2016}	0.814	0.103	p_{west}	0.490	0.075			
ψ_{2017}	0.813	0.106						
ψ_{2018}	0.741	0.128						

Table 15: Multiseason models of occupancy (ψ) for *L. tasmaniensis*. We present the AIC and distance from the most parsimonious model (ΔAIC).

ψ	γ	ϵ	p	AIC	ΔAIC
Int	Int	Int	Year	284.63	0.00
Int	Int	Int	Reg	286.40	1.77
Reg	Int	Int	Reg	288.46	3.82
Int	Year	Year	Year	289.76	5.13
Int	Int	Int	Int	289.78	5.15
Dist	Int	Int	Int	290.03	5.40
Reg	Int	Int	Int	291.95	7.32
Int	Year	Year	Int	293.81	9.18
Temp	Int	Int	Year	317.40	32.77
Dist	Int	Int	Temp	335.44	50.81
Reg	Int	Int	Temp	336.26	51.63

Table 16: Multiseason approach applied to the most parsimonious model from Table 15. Here we present the coefficient estimates for occupancy (ψ), colonisation (γ), extinction (ϵ) and detection (p) calculated with `colext` and their associated standard errors.

	EST	SE		EST	SE		EST	SE
ψ_{2014}	0.896	0.093	p_{2014}	0.418	0.077	ϵ	0.099	0.051
ψ_{2015}	0.827	0.113	p_{2015}	0.378	0.078	γ	0.0003	0.007
ψ_{2016}	0.813	0.118	p_{2016}	0.692	0.074			
ψ_{2017}	0.615	0.153	p_{2017}	0.576	0.112			
ψ_{2018}	0.585	0.168	p_{2018}	0.320	0.097			

B Heterogeneous Model Results

We considered several site and survey covariates however many of these were not collected with the accuracy we needed so had to be removed. Site covariates include distance to closest walking track (Dist) and region defined as geographic location of a group of sites within the sanctuary (Reg). Model selection was run in `occu` for each individual survey year and species. The most parsimonious models were then calculated using scripts supplied from Karavarsamis and Huggins (2018b) for the two-stage method.

Table 17: Single season model selection comparison for *C. signifera* with site and survey covariates. Covariates for ψ and p included distance to closest walking track (Dist), temperature (Temp), season and region (Reg) and present their associated z value, p-value and AIC to compare models.

Year	Method	ψ covariate	EST	SE	t	p	AIC
2014	Full	Int	2.29	0.99	2.32	0.02	57.62
	Two-stage	Int	2.29	0.99	2.31	0.02	57.62
2015	Full	Int	1.47	0.64	2.29	0.022	70.32
	Two-stage	Int	1.47	0.64	2.29	0.022	70.32
2016	Full	Int	2.72	1.04	2.61	0.009	62.19
	Two-stage	Int	2.72	1.04	2.61	0.009	62.19
2017	Full	Int	8.22	24.90	0.33	0.74	40.38
		Reg East	-6.83	24.90	-0.27	0.78	
		Reg West	-8.62	24.90	-0.35	0.73	
	Two-stage	DNC					
2018	Full	Int	1.48	0.65	2.29	0.022	73.05
	Two-stage	Int	1.48	0.65	2.29	0.022	73.05
Year	Method	p covariate	EST	SE	t	p	AIC
2014	Full	Int	2.63	1.04	2.52	0.11	57.62
		Reg East	-1.24	1.23	-1.01	0.31	
		Reg West	-2.57	1.21	-2.12	0.03	
	Two-stage	Int	2.63	1.04	2.53	0.01	57.62
		Reg East	-1.29	1.24	-1.05	0.30	
		Reg West	-2.57	1.26	-2.04	0.04	
2015	Full	Int	1.43	0.36	4.02	<0.001	70.32
	Two-stage	Int	1.43	0.36	4.02	<0.001	70.32
2016	Full	Int	1.73	0.36	4.77	<0.001	62.19
	Two-stage	Int	1.73	0.36	4.77	<0.001	62.19
2017	Full	Int	2.39	0.61	3.96	<0.001	40.38
	Two-stage	DNC					
2018	Full	Int	1.31	0.34	3.80	<0.001	73.05
	Two-stage	Int	1.31	0.34	3.80	<0.001	73.05

Table 18: Single season model selection comparison for *L. ewingii* with site and survey covariates. Covariates for ψ and p included distance to closest walking track (Dist), temperature (Temp), season and region (Reg) and present their associated z value, p-value and AIC to compare models.

Year	Method	ψ covariate	EST	SE	t	p	AIC
2014	Full	Int	0.89	0.65	1.36	0.17	59.97
	Two-stage	Int	0.93	0.69	1.35	0.18	60.55
2015	Full	Int	4.68	14.70	0.32	0.75	86.37
	Two-stage	Int	5.35	34.74	0.15	0.88	86.37
2016	Full	Int	9.3	44.5	0.21	0.83	82.74
	Two-stage	Int	1.88	1.48	1.27	0.20	89.70
2017	Full	Int	9.12	38.90	0.23	0.82	65.08
		Reg East	-7.49	39.00	-0.19	0.85	
		Reg West	-9.45	39.00	-0.24	0.81	
	Two-stage	DNC					
2018	Full	Int	0.74	0.71	1.03	0.30	72.75
	Two-stage	Int	0.87	0.72	1.20	0.23	125.78

Year	Method	p covariate	EST	SE	t	p	AIC
2014	Full	Int	-0.814	0.71	-1.14	0.25	59.97
		Reg East	0.95	0.88	1.07	0.28	
		Reg West	10.39	48.92	0.21	0.83	
	Two-stage	Int	-0.97	1.07	-0.90	0.37	60.55
		Reg East	0.72	1.26	0.57	0.57	
		Reg West	18.16	1339.42	0.01	0.99	
2015	Full	Int	-0.63	0.33	-1.90	0.06	86.37
	Two-stage	Int	-0.64	0.36	-1.76	0.08	86.37
2016	Full	Int	-1.10	0.47	-2.33	0.02	82.74
		Reg East	-1.10	0.65	1.69	0.09	
		Reg West	-0.29	0.73	-0.39	0.69	
	Two-stage	Int	-1.99	1.02	-1.96	0.05	89.70
		Reg East	1.81	1.14	1.59	0.11	
		Reg West	1.81	1.31	1.39	0.17	
2017	Full	Int	0.63	0.37	1.70	0.09	65.08
	Two-stage	DNC					
2018	Full	Int	-0.21	0.45	-0.46	0.65	72.75
		Dist Far	-1.40	0.93	-1.50	0.13	
	Two-stage	Int	-0.18	0.44	-0.40	0.69	125.78
		Dist Far	-15.86	1063.22	-0.15	0.99	

Table 19: Single season model selection comparison for *L. dumerilii* with site and survey covariates. Covariates for ψ and p included distance to closest walking track (Dist), temperature (Temp), season and region (Reg) and present their associated z value, p-value and AIC to compare models. DNC indicates that the model did not converge.

Year	Method	ψ covariate	EST	SE	t	p	AIC
2014	Full	Int	2.77	4.38	0.63	0.527	65.05
	Two-stage	Int	2.76	4.33	0.64	0.524	65.05
2015	Full	Int	8.51	39.1	0.22	0.828	62.73
	Two-stage	DNC					
2016	Full	Int	4.13	18.6	0.22	0.824	71.24
	Two-stage	Int	4.44	28.39	0.16	0.876	71.24
2017	Full	Int	2.29	4.19	0.55	0.584	57.93
	Two-stage	Int	2.29	4.19	0.55	0.584	57.93
2018	Full	DNC					
	Two-stage	DNC					
Year	Method	p covariate	EST	SE	t	p	AIC
2014	Full	Int	-0.60	0.50	-1.21	0.227	65.05
	Two-stage	Int	-0.60	0.50	-1.21	0.226	65.05
2015	Full	Int	-1.57	0.33	-4.74	<0.001	62.73
	Two-stage	DNC					
2016	Full	Int	-1.25	0.48	-2.60	0.009	71.24
	Two-stage	Int	-1.26	0.52	-2.42	0.016	71.24
2017	Full	Int	-0.97	0.62	-1.56	0.118	57.93
	Two-stage	Int	-0.97	0.62	-1.56	0.118	57.93
2018	Full	DNC					
	Two-stage	DNC					

Table 20: Single season model selection comparison for *L. tasmaniensis* with site and survey covariates. Covariates for ψ and p included distance to closest walking track (Dist), temperature (Temp), season and region (Reg) and present their associated z value, p-value and AIC to compare models. DNC indicates that the model did not converge.

Year	Method	ψ covariate	EST	SE	t	p	AIC
2014	Full	Int	8.61	38.4	0.22	0.823	67.51
	Two-stage	DNC					
2015	Full	Int	1.99	2.35	0.85	0.396	73.50
	Two-stage	Int	1.99	2.35	0.85	0.396	73.50
2016	Full	Int	3.60	2.68	1.34	0.18	90.04
	Two-stage	Int	3.49	2.36	1.48	0.14	208.38
2017	Full	Int	0.43	0.58	0.74	0.459	60.41
	Two-stage	Int	0.43	0.58	0.74	0.459	60.41
2018	Full	Int	7.97	46.60	0.17	0.864	55.98
	Two-stage	DNC					
Year	Method	p covariate	EST	SE	t	p	AIC
2014	Full	Int	-0.51	0.30	-1.71	0.087	67.51
	Two-stage	DNC					
2015	Full	Int	-1.01	0.47	-2.15	0.032	73.50
	Two-stage	Int	-1.01	0.47	-2.15	0.032	73.50
2016	Full	Int	0.25	0.35	0.72	0.47	90.04
		Dist Far	-1.35	0.67	-2010	0.05	
	Two-stage	Int	0.29	0.33	0.86	0.39	208.38
		Dist Far	-17.33	1240.81	-0.01	0.99	
2017	Full	Int	0.34	0.46	0.74	0.457	60.41
	Two-stage	Int	0.34	0.46	0.74	0.457	60.41
2018	Full	Int	-1.81	0.36	-5.03	<0.001	55.98
	Two-stage	DNC					