Biological Conservation 176 (2014) 99–108

Contents lists available at ScienceDirect

Biological Conservation

journal homepage: www.elsevier.com/locate/biocon

Location matters: Using spatially explicit occupancy models to predict the distribution of the highly mobile, endangered swift parrot

Matthew H. Webb [a,b,*], Simon Wotherspoon [c], Dejan Stojanovic [a], Robert Heinsohn [a], Ross Cunningham [a], Phil Bell [d], Aleks Terauds [a]

[a] Fenner School of Environment and Society, Australian National University, Canberra, ACT 0200, Australia
[b] Department of Primary Industries, Parks, Water and Environment, Hobart, Tasmania 7000, Australia
[c] Institute of Marine and Antarctic Studies, University of Tasmania, Sandy Bay, Tasmania 7005, Australia
[d] Department of Zoology, University of Tasmania, Sandy Bay, Tasmania 7005, Australia

ARTICLE INFO

Article history:
Received 23 November 2013
Received in revised form 5 May 2014
Accepted 8 May 2014
Available online 11 June 2014

Keywords:
Detectability
EM Algorithm
Spatial autocorrelation
Lathamus discolor
Zero-inflated Binomial models
Species distribution modelling

ABSTRACT

Occupancy modelling using data collected by repeatedly sampling sites is a common approach utilised by land managers to understand species distributions and trends. Two important factors that can complicate interpretation of these models are imperfect detection and spatial autocorrelation. We examine the effect of these potential errors using a multi-year data set on the distribution of the migratory and endangered swift parrot (*Lathamus discolor*). We simultaneously account for these effects by extending a zero-inflated Binomial (ZIB) framework to allow the inclusion of semiparametric, smooth spatial terms into both the occupancy and detection component of the model, in a maximum likelihood framework easily implemented in common software. This approach also has the advantage of relatively straightforward model selection procedures. We show that occupancy and detectability were strongly linked to food availability, but the strength of this relationship varied annually. Explicitly recognising spatial variability through the inclusion of semiparametric spatially smooth terms in the ZIBs significantly improved models in all years, and we suggest this predictor is an effective proxy for unmeasured environmental covariates or conspecific attraction. Importantly, the spatially explicit ZIBs predicted fewer occupied sites in more defined areas compared to non-spatial ZIBs. Given the importance of predicted distributions in land management, habitat protection and conservation of swift parrots, these models serve as an important tool in understanding and describing their ecology. Our results also reinforce the need for designing surveys that capture the underlying spatial structure of an ecosystem, especially when studying mobile aggregating species.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Effective population monitoring is fundamental to threatened species management and conservation planning (Martin et al., 2007; Sanderson et al., 2006). The importance of developing effective monitoring designs and analytical approaches has generated considerable discussion (Reynolds et al., 2011; Rhodes and Jonzén, 2011; Wintle et al., 2010), particularly regarding the need to identify and account for sources of error. When the results of monitoring identify the need for management responses that are contentious, expensive or impact on industry, accounting for error becomes especially important (Martin et al., 2007).

Highly mobile, rare or cryptic species can be difficult and expensive to monitor. Because resources are often limited, collecting detection/non-detection data from a sample of sites to be analysed within an occupancy-modelling framework is a popular approach among land management agencies (Kéry et al., 2013). Consequently, occupancy models and the relationship between occupancy and abundance, have been used extensively to estimate species density, distributions and habitat associations (e.g. Gaston et al., 2000; Hui et al., 2006). Estimating and accounting for false negative error rates or detection probability is fundamental to improving the reliability of occupancy models (MacKenzie et al., 2002; Martin et al., 2005; Royle and Nichols, 2003; Tyre et al., 2003; Wintle et al., 2004). The most common approach involves repeatedly sampling sites to estimate detection probability p, defined as the probability a species will be detected in a single site visit given that it occupies that site (MacKenzie et al., 2002). The

* Corresponding author. Address: DPIPWE, GPO Box 44, Hobart, Tasmania 7001, Australia. Tel.: +61 361654321.
E-mail address: Matthew.Webb@dpipwe.tas.gov.au (M.H. Webb).

http://dx.doi.org/10.1016/j.biocon.2014.05.017
0006-3207/© 2014 Elsevier Ltd. All rights reserved.

detection process is commonly influenced by the behaviour and abundance of the target species (Gu and Swihart, 2004; McCarthy et al., 2013), and if there is spatial or temporal heterogeneity in *p*, establishing its relationships with environmental variables can reduce bias in parameter estimators and improve sampling strategies (e.g. Bailey et al., 2004; Gibson, 2011; Lahoz-Monfort et al., 2014).

Errors in interpreting ecological relationships can also arise if spatial autocorrelation (SAC) is ignored, or not accounted for in the distribution of the target species (Dormann, 2007; Hawkins, 2012; Legendre, 1993). Generally, SAC originates from either an autocorrelated environment (i.e. where nearby locations are more similar than more distant ones) or through processes like conspecific attraction and limited dispersal ability of the target species (Lichstein et al., 2002). Importantly, recognition and analysis of SAC can provide insights into ecological processes that may otherwise be overlooked (Bini et al., 2009; Hawkins, 2012) and the effect of spatial structure has been recognised as an important component in modelling the occupancy-abundance relationship (Hui et al., 2006).

Recently, considerable attention has focused on improving statistical methods to account for either SAC or imperfect detection; however, relatively few studies have formally accounted for these processes simultaneously (but see Aing et al., 2011; Bled et al., 2011; Johnson et al., 2013; Royle et al., 2007). Some studies have accounted for spatial correlation in discrete spatial domains (Johnson et al., 2013; Royle et al., 2007; Wintle and Bardos, 2006), while others have focused on the detection process in transect based or cluster sampling designs (Aing et al., 2011; Guillera-Arroita et al., 2010, 2012; Hines et al., 2010). Other approaches model spatial variability through the inclusion of spatially correlated random fields (Diggle et al., 1998; Post van den Burg et al., 2011). Most of these studies, and other occupancy models that contain autocorrelation structure adopt a hierarchical Bayesian perspective (see also Gardner et al., 2010; Hoeting et al., 2000; Sargeant et al., 2005).

In this study, we use the endangered swift parrot (*Lathamus discolor*) to illustrate the importance of accounting for SAC and detection when modelling the distribution of mobile, cryptic and threatened species. Swift parrots are a migratory nectarivorous species seriously threatened by anthropogenic habitat loss throughout their range (Higgins, 1999). Their breeding range is restricted to the island of Tasmania, Australia, where they nest in tree hollows and rely on the erratic flowering of the Tasmanian blue gum (*Eucalyptus globulus* subsp. *globulus*) and black gum (*Eucalyptus ovata*) for food (Webb et al., 2012). However, there are few empirical data that quantify the relationship between nectarivores and flowering at macroecological scales. The very specific nesting and food requirements of the swift parrot, and the need for hollows and flowering to occur in the same area, make the species highly vulnerable to the effects of continuing habitat degradation and loss (Webb et al., 2012).

A key question for land managers is: how much habitat needs to be protected to conserve the species? Approximately one-third of the swift parrots potential breeding habitat is afforded varying levels of protection through the Comprehensive, Adequate and Representative (CAR) Reserve System (see Commonwealth of Australia, 1992). However, conservation (or protection) of non-reserved land (e.g. private land, production forest) that contains breeding habitat is highly contentious (Allchin et al., 2013) and can have serious economic implications for stakeholders. In this context, the relative importance of a particular area to swift parrots is often heavily scrutinised, especially where information is limited. Accurate, annual spatiotemporal information on the distribution of swift parrots, and the availability of their nesting and feeding habitat, is required to identify ecologically relevant spatial scales of management, prioritise key sites or regions, develop and inform off-reserve management actions, and set spatially explicit thresholds for habitat loss.

Given the dependence of swift parrots on flower for food, its use as a key explanatory variable was a logical starting point for our analyses. However, it is likely that other environmental or behavioural factors also influence occupancy and detection. From a logistical or economic perspective, it is often difficult to identify or measure these factors. We hypothesised that explicitly incorporating a smoothed spatial covariate in the occupancy and/or detectability component of zero-inflated Binomial models (ZIB) in a generalised additive model (GAM) framework, should help explain a large proportion of the variation due to these unknown or unmeasured factors. Our approach models the autocorrelation through smoothed functions of spatial coordinates where space is viewed as inherently continuous. This is in contrast to approaches that discretize space into regions or sites, and model spatial correlation through correlated random effects defined over sites in a Bayesian hierarchical framework (e.g. Bled et al., 2011; Johnson et al., 2013 and references therein; Wintle and Bardos, 2006). Our approach is more similar to geostatistical models in which spatial variability is modelled as spatially correlated random fields (e.g. Diggle et al., 1998; Post van den Burg et al., 2011). However, by modelling spatial variability through smooth functions of spatial coordinates rather than correlated random fields, our models can be fitted with standard maximum likelihood methods avoiding the need for complex Markov Chain Monte Carlo techniques.

Here we describe the design and implementation of a monitoring program, and associated analytical techniques, to better understand the spatial ecology of swift parrots and inform a landscape-scale conservation management strategy. We fitted Binomial models (with perfect detection), and zero-inflated Binomial models (that accounted for imperfect detection) with and without a smooth spatial covariate in GAM and generalised linear model (GLM) frameworks respectively, to test our hypothesis about the importance of spatial location. Using these models, we mapped the predicted distribution of swift parrots to illustrate dramatic spatiotemporal variation in their occurrence and detectability, while highlighting the importance of accounting for SAC. We also used simulated spatially structured data to form more generalised insights from our models.

2. Methods

2.1. Study area and sampling regime

We sampled across the known breeding range of the swift parrot (broadly defined by the natural range of *E. globulus*), which is restricted to Tasmania and covers approximately 10,000 km^2 (Fig. A1, Webb et al., 2012). Swift parrot detection/non-detection data were collected by repeatedly sampling a number of distinct sites over a three-week period in October 2009–2012 (number of sites ranged from 771 to 1034). A site was defined as a 200 m radius around a fixed point and the number of site visits, across all years, ranged from one to eight with a mean of 2.4 (see Appendix A for detailed sampling protocols). Flowering intensity (0–4 scale) was also recorded during these visits. Minimizing the amount of time taken for each annual survey reduced the likelihood of changes in detectability and violation of the assumption of closure, which is inherent in the models utilised (MacKenzie et al., 2006; Rota et al., 2009). A small subset of sites ($n = 16$) from the north-west of Tasmania that were geographically distinct from the rest of the sites (Fig. A1) was not used in the analyses to reduce their disproportionate impact as spatial outliers.

2.2. Spatial structure

The degree of SAC in the data was assessed using correlograms (based on Moran's I – Tiefelsdorf, 2000) of detection/non-detection (0, 1) and flowering score (0, 1, 2, 3, 4) for each year using Spatial Analyses in Macroecology v4 (SAM, Rangel et al., 2010).

2.3. Model form

While recognising that a range of factors potentially influence the probability of occupancy (Ψ) and probability of detection (p) of swift parrots, we deliberately kept the models simple, restricting the covariates to just flower and a semiparametric, smooth spatial term. Flower is recognised as a key driver of avian nectarivore distribution (Mac Nally and McGoldrick, 1997) and we hypothesised that other unmeasured environmental or behavioural factors would be captured by the spatial covariate. We also used the odds ratio of the flower coefficient to quantify the strength of its effect in the models.

Two classes of models were considered, a simple Binomial, where we assume p to be perfect and a ZIB, which accounts for imperfect detection. For the simple Binomial (Eq. (1)), y_i is a binary indicator that is 1 if the target species was detected on any visit to site i, and 0 otherwise.

$$Y_i \sim \text{Binomial}(1, \Psi_i) \tag{1}$$

Here Ψ_i is the probability that the target species is present at site i (assuming perfect detection), and is a function of the covariates.

The second class of models (ZIBs) assume that any site is either continuously occupied or unoccupied during the survey period, but detection is imperfect (but constant across the survey period). Hence, if the site is occupied, there is no guarantee the target species will be detected on any individual visit, and the observed detections are modelled with a zero-inflated Binomial distribution (Hall, 2000) (Eqs. (2a) and (2b)). In this case

$$Y_i \sim \text{Binomial}(n_i, z_i p_i) \tag{2a}$$

$$Z_i \sim \text{Binomial}(1, \Psi_i) \tag{2b}$$

where now y_i represents the number of times the target species was detected in n_i visits to the site. z_i is a latent binary variable that indicates whether a site is truly occupied. Here, $z_i = 1$ if the site is occupied and $z_i = 0$ if the site is unoccupied, and so $y_i > 0$ implies $z_i = 1$ and $z_i = 0$ implies $y_i = 0$. Again, Ψ_i is the probability that site i is occupied, and p_i is the conditional probability that a detection will occur on any single occasion if the site is occupied (assuming that detections occur independently). In turn, the probabilities of Ψ_i and p_i can be related to site-specific covariates.

2.4. Model fitting

Data were analysed for each year separately. First, we fitted the standard Binomial model using logistic regression with Ψ as a function of the variable *flower* (F) (Eq. (3)) using library glm in R (R Development Core Team, 2013). In vector notation

$$\log(\Psi/1 - \Psi) = \alpha + \beta_1 F \tag{3}$$

where Ψ is now the vector of probabilities that the target species is present at the sites, and F the vector of flower scores.

Autologistic models (Augustin et al., 1996) were then fitted, again with Ψ as a function of *flower* but with the addition of a spatial autocovariate (calculated using the default settings in SAM v4.0, logistic regression module) (Eq. (4)). Again, in vector notation

$$\log(\Psi/1 - \Psi) = \alpha + \beta_1 F + cWy \tag{4}$$

Here Wy represents the spatial autoregressive term, where y is now a binary vector indicating the presence or absence of the target species at each of the sites, W is the spatial relationship matrix that reflects the relation between each site and its neighbours, and c is the autoregressive parameter.

As a third alternative, we used the mgcv library in R (Wood, 2004) to fit a Binomial GAM that modelled Ψ as a function of *flower* and a smooth function $s(lat, lon)$ of location covariates (Eq. (5)). We allowed the mgcv package to select the appropriate level of smoothing.

$$\log(\Psi/1 - \Psi) = \alpha + \beta_1 F + s(lat, lon) \tag{5}$$

To examine the effect of adding a spatial covariate, we also compared the amount of SAC remaining in the residuals of each model fitted. AIC scores were used to rank the overall fit of the models (Burnham and Anderson, 2002).

Three approaches were used to fit the ZIBs. First, data were analysed separately for each year in program PRESENCE (Hines, 2012), where both Ψ and p are a function of *flower*. In principle, the ZIBs can also be fitted using VGAM package in R (Yee, 2010) and this package was used to fit ZIBs with *flower* as the only predictor. However, extensive tests on simulated data found VGAM ZIBs to be unreliable when a spatial covariate was added. Specifically, they would not converge with a spatial covariate term (i.e. the ZIB equivalent of Eq. (5)) and this approach was not pursued. Instead, ZIBs with the smooth location covariate were fitted using the EM Algorithm (Dempster et al., 1977) implemented in R (R Core Development Team, 2013). The EM Algorithm is an iterative procedure for deriving maximum likelihood estimates in the presence of missing data. The ZIB model can be fitted with the EM Algorithm by treating the latent z_i as missing data. Starting from an initial estimate, the EM Algorithm is an iterative two-step process that generates a sequence of estimates guaranteed to converge to the maximum likelihood estimate (see Appendix B for full details). While, circumstances exist where the ZIB model is degenerate, (e.g. with only one visit it is not possible to separate detection from occupancy); here we assume that sites are sampled sufficiently often that the likelihood has a unique maximum. To the best of our knowledge, this is the first successful implementation of ZIB occupancy models that incorporate a smoothed spatial covariate in a GAM framework using empirical ecological data. AICs were again used to compare the fit of all ZIBs, with the exception of the PRESENCE models, which were not comparable to other models due to differences in the way maximum likelihood is computed.

2.5. Predictions of occupancy and detectability

To compare predictions among all models, we generated spatially explicit estimates for Ψ (simple Binomial models and ZIBs) and p (ZIBs only). To better visualise the results, we interpolated the predictions at ecologically relevant scales (kriging with 0.02° pixel size, maximum of 50 neighbours and neighbour search radius of 0.05° (~5 km)) across the study area (implemented in Manifold Systems Professional V8 GIS software).

2.6. Simulations

Although we demonstrated that the EM Algorithm could be used to fit ZIBs, it was unclear whether these models were identifiable when the probabilities of Ψ and p are related to site specific covariates, or whether spatial variability in p will be confounded with variability in Ψ. If the two components cannot be clearly distinguished, there may be no value in fitting models that explicitly represent both p and Ψ. To address these issues, we first used simulated data to test if the model was indeed identifiable, and that

Table 1
Summary of the frequency of flower score (0, 1, 2, 3, 4) and the number of sites where swift parrots were detected, 2009–2012.

Flower score	2009	Sites birds detected	2010	Sites birds detected	2011	Sites birds detected	2012	Sites birds detected
0	586	16	373	28	787	35	621	29
1	70	3	149	29	83	20	135	16
2	61	16	138	35	85	33	103	32
3	45	31	152	54	46	23	100	54
4	9	6	38	13	33	26	27	21

variability in p could be distinguished from variability in Ψ. Second, we qualitatively compared the fit of the ZIBs with a presence-absence (or detection/non-detection) Binomial model (directly analogous to our simple GAMs).

We simulated data by generating N random sites distributed uniformly on the $[0, 1] \times [0, 1]$ square. Predictors were calculated for each site and used to construct the probability of occupancy at a site, and the probability of detecting the target species if the site is occupied. In these simulations, at least one visit was conducted at every site, and the number of additional visits was assumed to be Poisson distributed. Three of the predictors used were smooth functions of space, with two being sinusoids and the third a plane while the remaining two predictors were uniform random fields. Full details of the simulations are provided in Appendix C. The R code used to implement the EM Algorithm and the simulations is provided in Appendix D.

3. Results

3.1. Overview

The frequency of flower scores and sites where swift parrots were detected are summarised in Table 1. Naïve occupancy (i.e. proportion of sites where swift parrots were detected) over the four years ranged from 0.094 to 0.187. Flowering conditions varied between years (both in intensity and geographically), being generally very poor with localised flowering in 2009 (mainly in the south-eastern region), a mast flowering event in 2010 (again highest in the south–east), localised flowering in 2011(north-eastern region) and again in 2012 (mainly in the south) (Fig. 1).

3.2. Spatial structure

Correlograms indicated significant SAC in swift parrot detection/non-detection and *Eucalyptus* flowering in all years (Fig. 2a–d). Flowering was significantly spatially autocorrelated (Morans I test) out to a distance of 25–51 km and varied between years. In 2009, 2011 and a slightly lesser extent in 2012, the SAC in *detection* followed a similar pattern to that of *flower*, and was spatially autocorrelated out to approximately 30–35 km in those years; however, in 2010 *detection* was spatially autocorrelated out to a distance of over 40 km with a less similar pattern to that of *flower* (Fig. 2).

3.3. Models assuming perfect detectability ($p = 1$)

Not surprisingly, given the evidence for spatial structure in the data, all models improved (as indicated by lower AICs) when a spatial covariate was added. The best performing model in each year (where p was assumed to be perfect) was the Binomial GAM (Table 2, see also Table A1a–d for all model coefficients and odds ratios). Models in 2010 showed the most improvement after the addition of the spatial covariate, which reflected the reduced spatial dependency on flowering in that year. Odds ratios also showed that *flower* was much less important as a predictor in 2010 compared to other years (Table 2).

Correlograms of the residuals showed that the inclusion of the spatial predictor in the GAM removed all significant, positive SAC (Fig. A2a–d). In 2009 and 2011 (when flowering was more sparse but locally concentrated), *flower* alone accounted for far more of the spatial structure in the residuals compared to 2010 (Moran's I <0.1 in 2009 and 2011, and <0.2 in 2012 compared to 0.4 in 2010 in the first distance class, Fig. A2).

3.4. Zero-inflated Binomial models

Five ZIBs were fitted in each year (Tables 2 and A1). According to AICs, the VGAM and EM Algorithm models without spatial covariates were almost identical in their fit and all non-spatial ZIBs had very similar coefficients (Table A1). Models with a spatial covariate in either the Ψ and/or p component (i.e. GAM-ZIB framework) were better than those without the spatial covariate (i.e. GLM-ZIB framework) in all years (Table 2). The standard errors computed through the EM Algorithm are unreliable, and in general it is difficult to relate these to the true standard errors. Louis (1982) shows that the complete data information matrix required to compute the true standard errors can be expressed as the observed data information matrix adjusted for the information missing due to the missing observations (in our case, the true site occupancies). In principle, this result can be leveraged to estimate the true standard errors, but in practice this not a simple computation, and we could not see how to implement this in the general case. Therefore we have not included them in Table A1.

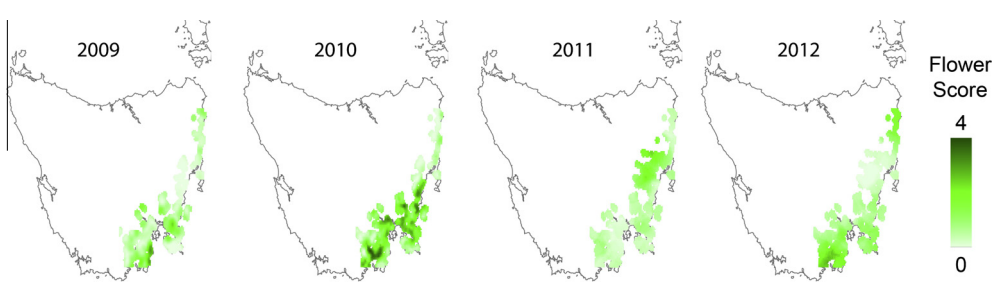

Fig. 1. Interpolated distribution of flowering over the swift parrot breeding range in each year. Flowering was scored on a 0–4 scale, and smoothed using kriging with a 0.02° cell size, 50 neighbours in a maximum 0.05° radius.

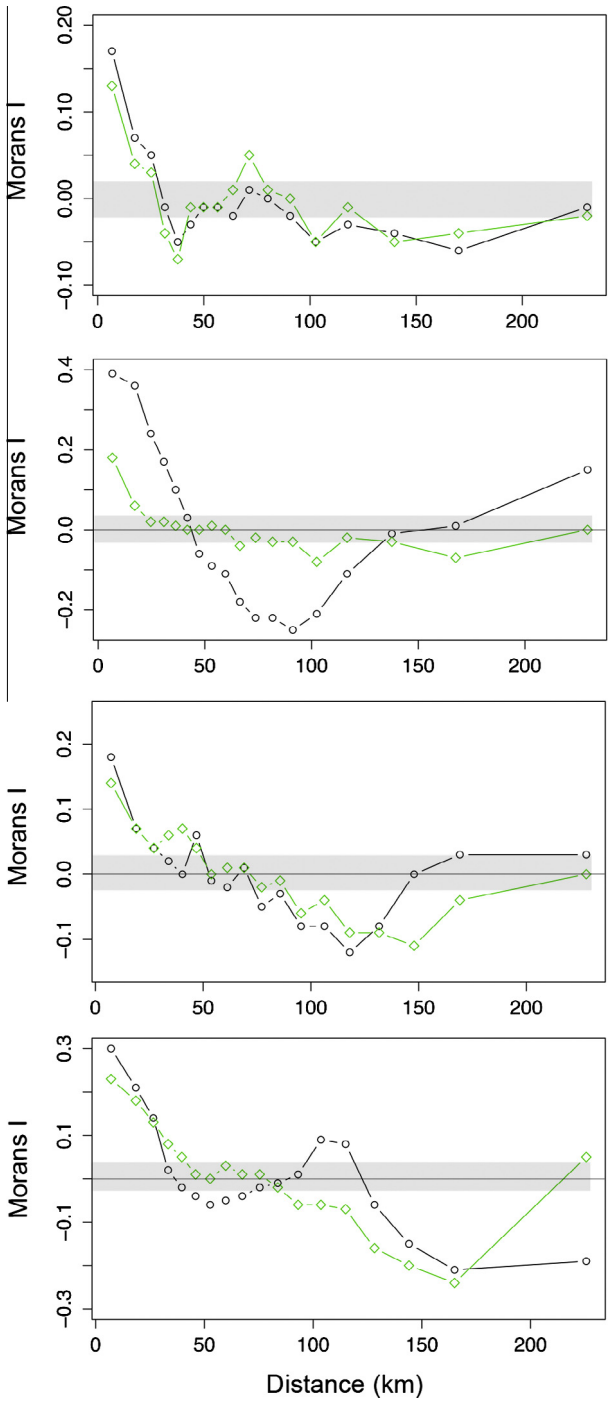

Fig. 2. Spatial autocorrelation (Morans I) in swift parrot detection/non-detection (black circles) and corresponding flower data (green diamonds) across the breeding range, for years 2009–2012. Shaded areas represent non-significant spatial autocorrelation ($p < 0.05$) as indicated by the Morans I test.

Odds ratios were again used to illustrate the relative importance of *flower* in the models. Odds ratios indicated that *flower* was an important predictor in modelling both Ψ *and* p in 2012, but had less influence on both components in 2010 (Table 2). By contrast, there were more obvious differences in the relative importance of *flower* between the Ψ and p components in 2009 (minimal influence on Ψ, strong effect on p in the two best models) and 2011 (very strong effect on Ψ, little predictive power for p across all models).

3.5. Model predictions of Ψ and p

Predictions from the best simple GAMs (where $p = 1$) all showed high probability of swift parrot presence in fairly discrete kernels (Fig. 3a). These areas differed among years, suggesting that swift parrots are not only flexible in their habitat utilisation, but also utilised much of the available habitat over time. While these predictions are useful for comparative purposes, from here on we focus more on models that account for imperfect detection.

The inclusion of the spatial predictor in the p component of the ZIB models highlighted considerably more heterogeneity in the detection process than that observed in the models without spatial covariates (mean range of p in spatial models = 0–0.89 *cf.* 0.22–0.64 in non-spatial models, Fig. A3, A4a and b). The low estimates of p in the spatial models were generally at sites geographically disjunct from clusters of sites where birds were detected (i.e. the informative sites for estimating p). This also resulted in markedly different Ψ predictions, with generally high probabilities in a relatively narrow range (mean range of Ψ: 0.58–0.99; Fig. A4c). The high predictions over the narrow range can largely be attributed to uncertainty surrounding estimates at many sites where p was very low, which in turn is likely related to the number of visits (see also Appendix C – p6 and Fig. 13). Low detectability also confounded realistic occupancy predictions in the models that included a spatial covariate in both components (Fig. A5a).

Although AIC scores suggested that models with the spatial component in detection often performed better (Table 2), the resulting Ψ predictions were clearly unsatisfactory (Figs. A4c, A5a). For this reason we focus on Ψ estimates from ZIB models with the spatial covariate only in the Ψ component. Non-spatial ZIBs showed more diffuse predictions (Fig. 3b), and the range of predictions was typically smaller than those models with a spatial predictor in the Ψ component (mean Ψ range: 0.13–0.80 *cf.* 0–0.97, Fig. A6). Direct comparisons of the predictions indicate non-spatial models (Fig. 3b) generally over and underestimate Ψ compared to models with a spatial covariate (Fig. 3c).

ZIBs with the spatial covariate in the Ψ component provided similar predictions to the simple Binomial GAMs, with concentrated kernels of high probability and areas of low probability over much of the breeding range (Fig. 3a and c). Despite the apparent visual similarities of the predictions from the two models, site-by-site comparisons clearly highlighted the influence of p on Ψ predictions (Fig. A7). For example, in 2010 there was close agreement between Ψ estimates (Fig. A7), with relatively constant p across the range (Fig. A8). By contrast, in 2009 there was much less agreement in predictions between the two models (Fig. A7). This year differed from others in that a relatively high proportion of sites were only visited once (~50%), and these had very low estimates of p (median < 0.1 – Figs. A3a, A8).

3.6. Simulations

The simulations showed that when p was constant, both the zero-inflated and simple Binomial models detect the factors influencing Ψ, including spatial structure (Appendix C). However, consistent with the empirical results reported above, when p varies, the ZIB is able to separate factors influencing p and Ψ (Fig. 4). By contrast, the simple GAMs (where $p = 1$) confound the factors influencing p with those influencing Ψ and conflate these estimates (Fig. 4, Appendix C). For example, if we take several forms of clear spatial structure (Fig. 4a) and incorporate them into our simulated models, the linear trend (f_3 in Fig. 4a) in p across the domain is not reflected in the fitted smooth of the Ψ component of the ZIB (Fig. 4b). However, estimates of Ψ from the simple GAM suggest that the latter both overestimates and underestimates Ψ across much of the space (Fig. 4c). Similarly, in Fig. 4d, the spatial trends

Table 2
Form of models fitted and corresponding AIC values where Ψ = probability of occurrence, p = probability of detection and bracketed terms represent the covariates included in the models. *Flower* = score 1–4; $s(lat, lon)$ = bivariate smooth location term. Note that the AICs of the simple models and the AICs of the zero-inflated Binomial models (ZIBs) are not comparable; AICs for PRESENCE model are not comparable with those fitted with the EM Algorithm and AICs are not comparable across years.

Model	Implementation	AIC (2009)	AIC (2010)	AIC (2011)	AIC (2012)
$\Psi(flower)\cdot p(1)$	SAM[a]	288	756	571	631.5
$\Psi(flower + cW)\cdot p(1)$	SAM[a]	251	509	458	488
$\Psi(flower + s(lat, lon)\cdot p(1)$	R-package: mgcv[b]	237 (5.0)[f]	507 (1.6)[f]	447 (3.0)[f]	473 (2.8)[f]
Zero-inflated Binomial models					
$\Psi(flower)\cdot p(flower)$	PRESENCE[c]	689	1306	1006	1003
$\Psi(flower)\cdot p(flower)$	R-package: VGAM[d]	Did not converge	1035	782	832
$\Psi(flower)\cdot p(flower)$	EM Algorithm[e]	448 (2.8, 2.5)	1035 (1.8, 1.1)	782 (3.2, 1.4)	832 (2.2, 2.0)
$\Psi(flower)\cdot p(flower + s(lat, lon))$	EM Algorithm[e]	400 (1.5, 4.2)	807 (1.5, 1.4)	672 (5.9, 1.5)	683[g] (3.4, 2.2)
$\Psi(flower + s(lat, lon))\cdot p(flower)$	EM Algorithm[e]	440 (2.6, 2.4)	808 (1.5, 1.1)	665[g] (6.7, 1.3)	694 (2.0, 1.9)
$\Psi(flower + s(lat, lon))\cdot p(flower + s(lat, lon))$	EM Algorithm[e]	390[g] (1.9, 4.5)	783[g] (2.0, 1.4)	671 (7.2, 1.3)	686 (2.5, 2.4)

[a] Spatial Ecology in Macroecology – Rangel et al. (2010).
[b] Wood (2004).
[c] Hines (2012).
[d] Yee (2010).
[e] Implementation developed in this study (see Appendix B for details).
[f] Denotes best models for $p = 1$.
[g] Denotes best ZIB. Odds ratios indicating the relative strength of the variable *flower* are shown in brackets after the AIC values.

in both Ψ and p are distinguished much more accurately in the ZIB and while Ψ estimates from the simple GAM do capture some of the spatial structure, they are clearly confounded by the factors influencing p as well. For more details on the model simulations and associated results see Appendix C.

4. Discussion

Our study demonstrates that complex and spatiotemporally variable interactions between a difficult to study species and their food source can be effectively monitored and modelled to inform conservation management. Through the use of a smoothed spatial covariate in occupancy models, we provide another tool that can be used to address common challenges facing survey design and associated distributional analyses that typically hamper monitoring and conservation efforts for mobile, aggregating animal populations. Untangling the processes influencing detection and occupancy is important for avoiding misleading inferences, especially as such processes are often inextricably linked, with some clearly influencing both parameters while others may only affect one or the other. We highlight the importance of accounting for spatial autocorrelation, not only for modelling animal occurrence, but also for understanding the detection process. Furthermore, the potential of more flexible models (such as the GAM based models we utilise here) has been recognised, and while it has been suggested that they may produce superior occupancy models, they had yet to be implemented successfully with empirical data (Martin and Fahrig, 2012).

Across all models, the importance of *flower* as a predictor increased as its availability decreased (e.g. 2009). By contrast, the importance of site location (i.e. the smoothed spatial covariate) increased with the availability of flowering (e.g. 2010), suggesting that other processes also influenced swift parrot distribution. The empirical observations were supported by the simulations, which demonstrated that incorporating the spatial covariate into ZIBs allowed the spatial structure present in both Ψ and p to be identified, whereas the simple Binomial GAMs confound the factors influencing Ψ and p. When spatial structure is present (as is typically the case for mobile, aggregating species), ignoring spatial location in the absence of other explanatory variables when modelling Ψ or p can mislead inferences. In the swift parrot models the importance of the spatial covariate varied between Ψ and p in each year. However, the inclusion of the spatial covariate in the p component of the ZIB produced very low estimates (e.g. <0.01) for sites that were geographically distinct from the informative sites (i.e. sites where birds were detected), which in turn resulted in unreliable estimates of Ψ for those sites. Typically this can be attributed to too few repeated visits (e.g. Guillera-Arroita et al., 2010), but can also be related to the type of model fitted (and the interactions between the occupancy and detectability component).

Notwithstanding these limitations, the spatially explicit estimates of detectability may provide important insights into interpreting spatial variation in swift parrot population density. When Ψ and p both increase with a covariate, as they do in our study with *flower*, this can indicate the abundance of the target species is responding to the covariate (Yackulic et al., 2013). Such a trend would suggest a positive abundance–occupancy and/or abundance–detectability relationship (Gaston et al., 2000; McCarthy et al., 2013). We argue that much of the heterogeneity in both Ψ and p in this study originates from variations in the abundance of swift parrots over multiple spatial scales (i.e. from site to landscape scales). Here, the detection process is likely influenced by (i) increased calling frequency as abundance increases at the site level, and (ii) increased abundance of birds in the landscape surrounding a site, increasing the probability of a bird being present and therefore detected at a site when it is sampled. These complex interactions highlight the importance of studying the ecological mechanisms driving the occupancy/detection processes in spatially structured systems, and understanding the response of aggregated species distributions to the influence of environmental drivers at different spatial scales (Hui et al., 2010; Martin and Fahrig, 2012; Welsh et al., 2013).

The importance of testing and accounting for SAC will vary depending on the scale and level of aggregation of the target species and environmental predictors. The advantage of our approach (i.e. smoothed spatial location) is that it provides a more natural description of spatial structuring (or aggregation) because there is no requirement for *a priori* groupings of sites into clusters, transects or discrete spatial domains (e.g. Aing et al., 2011; Hines et al., 2010; Johnson et al., 2013), which can be arbitrary or ecologically irrelevant (Guillera-Arroita et al., 2011). Our methods also provide the potential to account for more sources of heterogeneity in Ψ or p, and improve understanding of bias in parameter estimators. Our approach is likely to be particularly relevant to surveys conducted at large spatial scales in dynamic systems when few ecologically relevant covariates are available, or when the scale of effect of an environmental factor is unknown or varies across multiple temporal and spatial scales. For example, in this study *flower* is important at the site level (i.e. 200 m radius), but our predictions also suggest

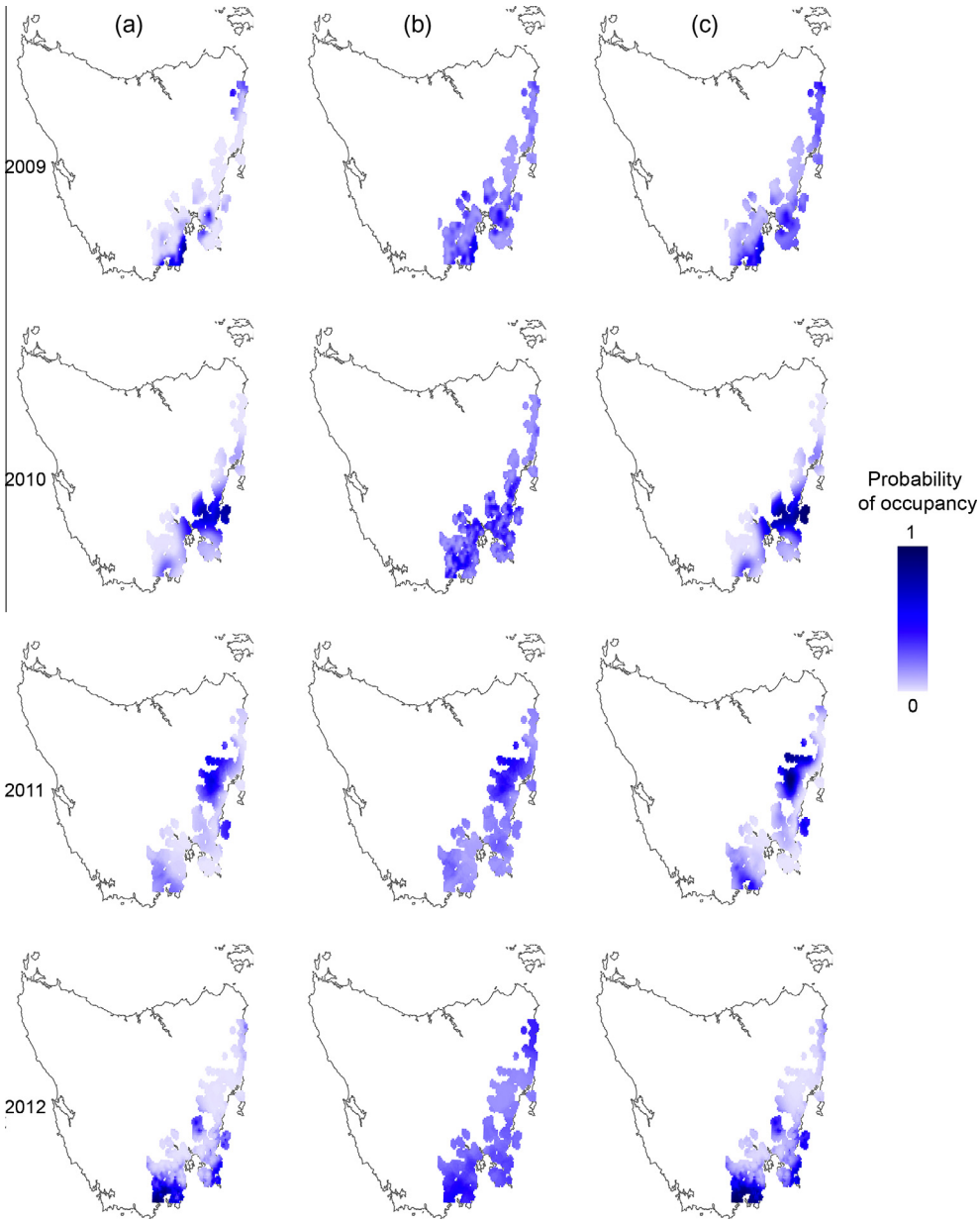

Fig. 3. Prediction maps of swift parrot occupancy 2009–2012, showing: (a) simple generalised additive models; (b) standard zero-inflated Binomial models with generalised linear models and no spatial covariate; and (c) zero-inflated Binomial models with the spatial covariate in the occupancy component only. Predictions are smoothed from point data using kriging with a 0.02° cell size, 50 neighbours in a maximum 0.05° radius.

its effect operates at much larger spatial scales. Further, the scales of effect vary from year to year depending on overall flowering conditions.

While the geostatistical methods mentioned above also view space as continuous, our approach provides an alternative with the advantage of much simpler model selection procedures. We hope this ease of implementation and model assessment makes our method more accessible to land managers, which may not always have the statistical knowledge to tackle complex problems or the resources to implement more complex analytical techniques. While our implementation using the EM Algorithm has the disadvantage of not providing standard errors around estimates, the improvement in the models, and consequent improvement of occupancy estimates, identify discrete defined areas of important habitat at ecologically relevant scales. On the whole, we think it better to utilise the advantages of more flexible, GAM based ZIBs (as noted by Martin and Fahrig, 2012), and suggest that in many cases, a better model with no standard errors is more informative than a weaker model with standard errors. While the EM Algorithm implementation represents a compromise in this respect, it is likely that future work will develop implementations that include estimates of uncertainty.

Often budgetary and logistic constraints may prevent adequate spatial replication to capture spatial structure and undertake multiple repeat visits. However, we show that devising a sampling design that captures the underlying spatial structure of the study system can be just as important as addressing imperfect detection. Additionally, as recently noted by Johnson et al. (2013), exploiting SAC may reduce the need for many repeat visits because spatial dependence between sites compensates for the lack of temporal replication. However, it is a trade-off, and we recognise that repeat surveys will always decrease the likelihood of recording false

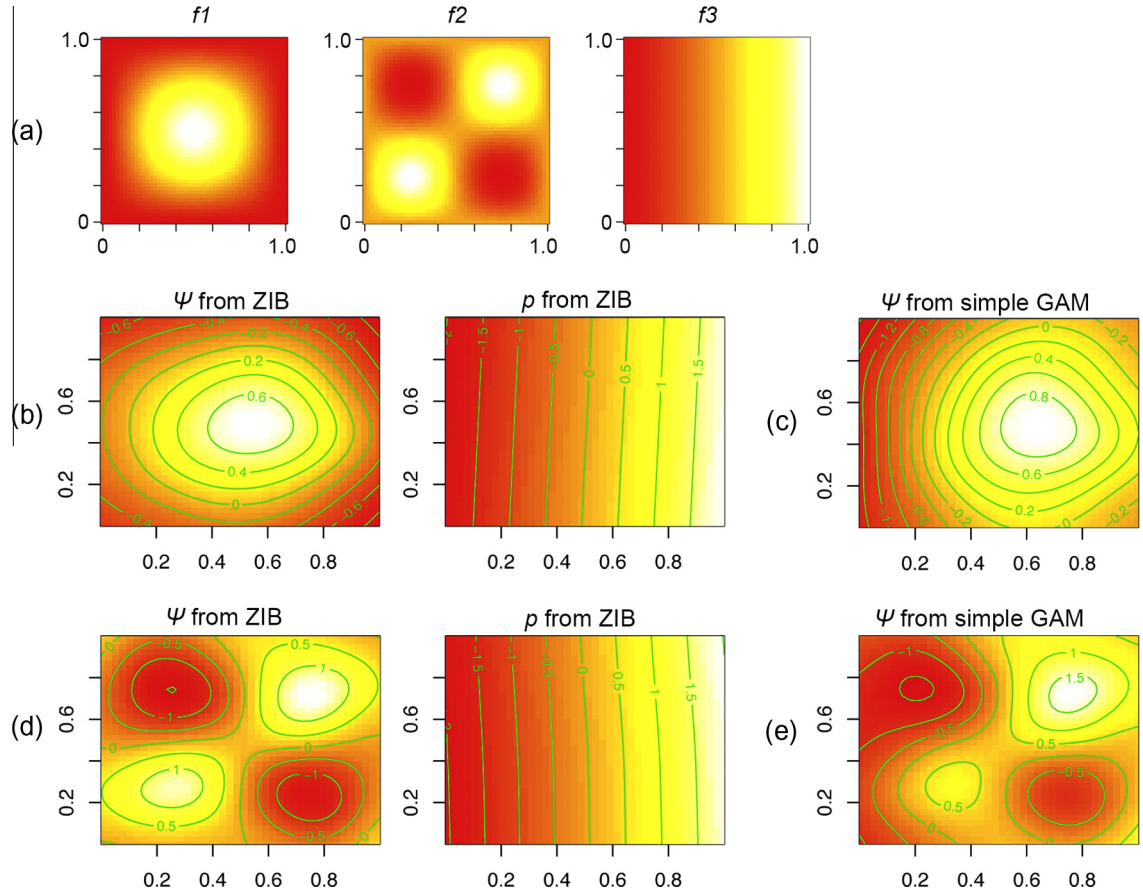

Fig. 4. Fitted smooths from the GAM simulations when detectability varies across the domain: (a) the structure of the three spatially smooth predictors; (b) zero-inflated Binomial model (ZIB) - occupancy (Ψ) spatial structure from f_1 and detectability (p) spatial structure from f_3; (c) simple Binomial generalised additive mode (GAM) – occupancy from model including f_1 and f_3; (d) zero-inflated Binomial model – Ψ spatial structure from f_2 and p spatial structure from f_3; and (e) simple Binomial GAM – occupancy from model including f_2 and f_3 (see Appendix C for further details and all simulated model comparisons).

absences. This study, together with other recent reviews, have highlighted the value of considering and utilising multiple statistical frameworks to better understand underlying ecological mechanisms, and to avoid misleading inferences (Martin and Fahrig, 2012; Welsh et al., 2013).

5. Implications for swift parrot conservation

Our study describes dramatic spatiotemporal variation in the swift parrot population driven by the distribution of *Eucalyptus* flowering. To our knowledge, this is the first population level study of a highly mobile nectarivore to describe macroecological patterns in distribution and demonstrate a spatial dependency on flowering. Understanding the variation in importance of the spatial covariate may provide insights into the mechanisms driving variations in the abundance of swift parrots. The SAC not explained by *flower* may be due to processes such as conspecific attraction or the omission of other unmeasured but ecologically important explanatory variables (e.g. availability of nesting sites or variations in the density of food trees). Similarly, the spatial covariate may explain larger scale effects, whereby site quality (as perceived by swift parrots) may also be dependent on flowering conditions at larger spatial scales than our site level measurement.

The spatially explicit models developed here represent a significant improvement on non-spatial models (as indicated by AICs) and the associated predictions of Ψ are over much smaller and more discrete areas. Improving models and associated predictions in this way not only reduces uncertainty about the species distribution, but also provides land managers with more confidence in making decisions that affect other stakeholders. Our findings are critical to informing these decisions because: (i) only a fraction of the breeding range appears suitable (and occupied) in most years (due to the co-occurrence of hollows and flowering), but that fraction varies considerably between years; (ii) we can identify focal regions for protection or restoration (see Fig. 3c), and provide robust quantitative thresholds (i.e. occupancy probability) on which to base these decisions; (iii) they allow an assessment of habitat availability to better inform the development of spatially explicit off-reserve conservation strategies and; (iv) we provide an analytical framework for understanding population level processes into the future (e.g. predation risk – see Stojanovic et al., 2014) and the likely impacts of climate change on flowering. Most importantly, we clearly identify areas of the breeding range that need to be managed in a way that provides enough habitat for the majority of the population to breed and forage in a given year.

Our study was developed to tackle the challenges inherent in devising an effective and informed conservation strategy for the endangered swift parrot. Many other species behave in similar ways, but due to statistical and logistic challenges, rigorous monitoring of their movements and ecology across multiple spatial scales has not been possible (Newton, 2006). Our methods represent an effective tool for monitoring and modelling these difficult to study species, and form the basis of an accessible analytical framework to assist conservation managers in overcoming the barriers to collecting informative and reliable distributional data.

Acknowledgments

This study was largely funded by the Australian Government through a Caring for Our Country Grant. Fiona Hume, Dave James, Shaun Thurstans, Alan Wiltshire, Chris Spencer, Chris Tzaros and over 35 skilled volunteers from BirdLife Australia provided invaluable support during data collection. The Department of Primary Industries, Parks, Water and Environment and the Forest Practices Authority provided further logistical support. Melodie McGeoch and three anonymous reviewers provided useful comments that improved the manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.biocon.2014.05.017.

References

Aing, C., Halls, S., Oken, K., Dobrow, R., Fieberg, J., 2011. A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites. J. Appl. Ecol. 48, 1508–1517.
Allchin, R., Kirkpatrick, J., Kriwoken, L., 2013. On not protecting the parrot: impact of conservation and planning legislation on an endangered species in Tasmania. J. Wildlife Law Policy 16, 81–104.
Augustin, N.H., Mugglestone, M.A., Buckland, S.T., 1996. An autologistic model for the spatial distribution of wildlife. J. Appl. Ecol. 33, 339–347.
Bailey, L.L., Simons, T.R., Pollock, K.H., 2004. Estimating site occupancy and species detection probability parameters for terrestrial salamanders. Ecol. Appl. 14, 692–702.
Bini, L.M., Diniz, J.A.F., Rangel, T., Akre, T.S.B., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araujo, M.B., Baselga, A., Beck, J., Bellocq, M.I., Bohning-Gaese, K., Borges, P.A.V., Castro-Parga, I., Chey, V.K., Chown, S.L., de Marco, P., Dobkin, D.S., Ferrer-Castan, D., Field, R., Filloy, J., Fleishman, E., Gomez, J.F., Hortal, J., Iverson, J.B., Kerr, J.T., Kissling, W.D., Kitching, I.J., Leon-Cortes, J.L., Lobo, J.M., Montoya, D., Morales-Castilla, I., Moreno, J.C., Oberdorff, T., Olalla-Tarraga, M.A., Pausas, J.G., Qian, H., Rahbek, C., Rodriguez, M.A., Rueda, M., Ruggiero, A., Sackmann, P., Sanders, N.J., Terribile, L.C., Vetaas, O.R., Hawkins, B.A., 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography 32, 193–204.
Bled, F., Royle, J.A., Cam, E., 2011. Hierarchical modelling of an invasive spread: the Eurasian Collared-Dove *Streptopelia decaocto* in the United States. Ecol. Appl. 21, 290–302.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multi-Model Inference, second ed. Springer Verlag, New York.
Commonwealth of Australia, 1992. Joint ANZECC/MCFFA National Forest Policy Statement Implementation sub-committee (JANIS), Nationally Agreed Criteria for the establishment of a Comprehensive, Adequate and Representative Reserve System for forests in Australia, Canberra.
Dempster, A., Laird, N., Rdin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. B 39, 1–38.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. Roy. Stat. Soc. 47, 299–350.
Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Glob. Ecol. Biogeogr. 16, 129–138.
Gardner, C.L., Lawler, J.P., Hoef, J.M.V., Magoun, A.J., 2010. Coarse-scale distribution surveys and occurrence probability modeling for wolverine in interior Alaska. J. Wildlife Manage. 74, 1894–1903.
Gaston, K.J., Blackburn, T.M., Greenwood, J.J.D., Gregory, R.D., Quinn, R.M., Lawton, J.H., 2000. Abundance–occupancy relationships. J. Appl. Ecol. 37, 39–59.
Gibson, L.A., 2011. The importance of incorporating imperfect detection in biodiversity assessments: a case study of small mammals in an Australian region. Divers. Distrib. 17, 613–623.
Gu, W., Swihart, R.K., 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. Biol. Conserv. 116, 195–203.
Guillera-Arroita, G., Ridout, M.S., Morgan, B.J.T., 2010. Design of occupancy studies with imperfect detection. Methods Ecol. Evol. 1, 131–139.
Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S., Linkie, M., 2011. Species occupancy modelling for detection data collected along a transect. J. Agri. Biol. Environ. Stat. 3, 301–317.
Guillera-Arroita, G., Ridout, M.S., Morgan, B.J.T., Linkie, M., 2012. Models for species-detection data collected along transects in the presence of abundance-induced heterogeneity and clustering in the detection process. Methods Ecol. Evol. 3, 358–367.
Hall, D.B., 2000. Zero-inflated Poisson and Binomial regression with random effects: a case study. Biometrics 56, 1030–1039.
Hawkins, B.A., 2012. Eight (and a half) deadly sins of spatial analysis. J. Biogeogr. 39, 1–9.
Higgins, P.J. (Ed.), 1999. Handbook of Australian, New Zealand and Antarctic Birds. Oxford University Press, Melbourne.
Hines, J.E. 2012. PRESENCE 5.5 – Software to estimate patch occupancy and related parameters. USGS-PWRC. Available at: <http://www.mbr-pwrc.usgs.gov/software/presence.html>.
Hines, J.E., Nichols, J.D., Royle, J.A., MacKenzie, D.I., Gopalaswamy, A.M., Kumar, N.S., Karanth, K.U., 2010. Tigers on trails: occupancy modeling for cluster sampling. Ecol. Appl. 20, 1456–1466.
Hoeting, J.A., Leecaster, M., Bowden, D., 2000. An improved model for spatially correlated binary responses. J. Agri. Biol. Environ. Stat. 5, 102–114.
Hui, C., McGeoch, M.A., Warren, M., 2006. A spatially explicit approach to estimating species occupancy and spatial correlation. J. Anim. Ecol. 75, 140–147.
Hui, C., Veldtman, R., McGeoch, M., 2010. Measures, perceptions and scaling patterns of aggregated species distributions. Ecography 33, 95–102.
Johnson, D.S., Donn, P.B., Hooten, M.B., Ray, J.C., Pond, B.A., 2013. Spatial occupancy models for large data sets. Ecology 94, 801–808.
Kéry, M., Guillera-Arroita, G., Lahoz-Monfort, J.J., 2013. Analysing and mapping species range dynamics using occupancy models. J. Biogeogr. 40, 1463–1474.
Lahoz-Monfort, J.J., Guillera-Arroita, G., Wintle, B.A., 2014. Imperfect detection impacts the performance of species distribution models. Glob. Ecol. Biogeogr. 23, 504–515.
Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? Ecology 74, 1659–1673.
Lichstein, J.W., Simons, T.R., Shriner, S.A., Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. Ecol. Monogr. 72, 445–463.
Louis, T.A., 1982. Finding the observed information when using the EM Algorithm. J. Roy. Stat. Soc. B44, 226–233.
Mac Nally, R., McGoldrick, J.M., 1997. Landscape dynamics of bird communities in relation to mass flowering in some eucalypt forests of central Victoria, Australia. J. Avian Biol. 28, 171–183.
MacKenzie, D., Nichols, J., Lachman, G., Droege, S., Royle, J.A., Langtimm, C., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255.
MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2006. Occupancy Estimation and Modelling: Inferring Patterns and Dynamics of Species Occurrence. Elsevier, San Diego, California.
Martin, A.E., Fahrig, L., 2012. Measuring and selecting scales of effect for landscape predictors in species-habitat models. Ecol. Appl. 22, 2277–2292.
Martin, T., Wintle, B., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J.H., Possingham, H.P., 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. 8, 1235–1246.
Martin, J., Kitchens, W.M., Hines, J.E., 2007. Importance of well-designed monitoring programs for the conservation of endangered species: case study of the snail kite. Conserv. Biol. 21, 472–481.
McCarthy, M.A., Moore, J.L., Morris, W.K., Parris, K.M., Garrard, G.E., Vesk, P.A., Rumpff, L., Giljohann, K.M., Camac, J.S., Bau, S.S., Friend, T., Harrison, B., Yue, B., 2013. The influence of abundance on detectability. Oikos 122, 717–726.
Newton, I., 2006. Advances in the study of irruptive migration. Ardea 94, 433–460.
Post van den Burg, M., Bly, B., VerCauteren, T., Tyre, A., 2011. Making better sense of monitoring data from low density species using a spatially explicit modeling approach. J. Appl. Ecol. 48, 47–55.
R Development Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Rangel, T.F., Diniz-Filho, J.A.F., Bini, L.M., 2010. SAM: a comprehensive application for spatial analysis in macroecology. Ecography 33, 46–50.
Reynolds, J.H., Thompson, W.L., Russell, B., 2011. Planning for success: identifying effective and efficient survey designs for monitoring. Biol. Conserv. 144, 1278–1284.
Rhodes, J.R., Jonzén, N., 2011. Monitoring temporal trends in spatially structured populations: how should sampling effort be allocated between space and time? Ecography 34, 1040–1048.
Rota, C.T., Fletcher Jr., R.J., Dorazio, R.M., Betts, M.G., 2009. Occupancy estimation and the closure assumption. J. Appl. Ecol. 46, 1173–1181.
Royle, J.A., Nichols, J.D., 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84, 777–790.
Royle, J.A., Kéry, M., Gautier, R., Schmid, H., 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. Ecol. Monogr. 77, 465–481.
Sanderson, F.J., Donald, P.F., Pain, D.J., Burfield, I.J., van Bommel, F.P.J., 2006. Long-term population declines in Afro-Palearctic migrant birds. Biol. Conserv. 131, 93–105.
Sargeant, G., Solvada, M., Slivinski, C., Johnson, D., 2005. Markov chain Monte Carlo estimation of species distributions: a case study of the swift fox in western Kansas. J. Wildlife Manage. 69, 483–497.
Stojanovic, D., Webb, M.H., Alderman, R., Porfirio, L.L., Heinsohn, R., 2014. Discovery of a novel predator reveals extreme but highly variable mortality for an endangered migratory bird. Divers. Distrib. http://dx.doi.org/10.1111/ddi.12214.
Tiefelsdorf, M., 2000. Modeling Spatial Processes: the Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Morans' I. Cambridge University Press, Cambridge.
Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K., Possingham, H.P., 2003. Improving precision and reducing bias in biological surveys: estimating false negative error rates. Ecol. Appl. 13, 1790–1801.

Webb, M.H., Holdsworth, M.C., Webb, J., 2012. Nesting requirements of the endangered Swift Parrot (*Lathamus discolor*). Emu 112, 181–188.
Welsh, A.H., Lindenmayer, D.B., Donnelly, C.F., 2013. Fitting and interpreting occupancy models. PLoS ONE 8, e52015.
Wintle, B.A., Bardos, D.C., 2006. Modeling species-habitat relationships with spatially autocorrelated observation data. Ecol. Appl. 16, 1945–1958.
Wintle, B.A., McCarthy, M.A., Parris, K.M., Burgman, M.A., 2004. Precision and bias of methods for estimating point survey detection probabilities. Ecol. Appl. 14, 703–712.
Wintle, B.A., Runge, M.C., Bekessy, S.A., 2010. Allocating monitoring effort in the face of unknown unknowns. Ecol. Lett. 13, 1325–1337.
Wood, S.N., 2004. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J. Am. Stat. Assoc. 99, 673–686.
Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H., Veran, S., 2013. Presence-only modelling using MAXENT: when can we trust the inferences? Methods Ecol. Evol. 4, 236–243.
Yee, T., 2010. The VGAM package for categorical data analysis. J. Stat. Softw. 32, 1–34.