Climate and the range dynamics of species with imperfect detection

Res Altwegg[1,2,*]**, Marius Wheeler**[1]
and Birgit Erni[3]

[1]*Animal Demography Unit, Department of Zoology, and*
[3]*Department of Statistical Sciences, University of Cape Town, Rondebosch 7701, Republic of South Africa,*
[2]*Global Change & Biodiversity Programme, South African National Biodiversity Institute, Private Bag X7, Claremont 7735, Republic of South Africa*
Author for correspondence (res.altwegg@adu.org.za).

Reliable predictions for species range changes require a mechanistic understanding of range dynamics in relation to environmental variation. One obstacle is that most current models are static and confound occurrence with the probability of detecting a species if it occurs at a site. Here we draw attention to recently developed occupancy models, which can be used to examine colonization and local extinction or changes in occupancy over time. These models further account for detection probabilities, which are likely to vary spatially and temporally in many datasets. Occupancy models require repeated presence/absence surveys, for example checklists used in bird atlas projects. As an example, we examine the recent range expansion of hadeda ibises (*Bostrychia hagedash*) in South African protected areas. Colonization exceeded local extinction in most biomes, and the probability of occurrence was related to local climate. Extensions of the basic occupancy models can estimate abundance or species richness. Occupancy models are an appealing additional tool for studying species' responses to global change.

Keywords: *Bostrychia hagedash*; colonization; detection probability; extinction; global change; occupancy model

1. INTRODUCTION

Predicted changes in biodiversity due to global change are mainly based on static species distribution models relating observed species occurrence to climate (Thomas *et al.* 2004; Huntley *et al.* 2006). An essential next step is to quantify range changes, local extinctions and colonizations, and relate them to environmental variation. This requires data on species occurrence at a fine spatio-temporal scale. At this scale, a major difficulty for understanding geographical distributions of animals is that species are not detected everywhere they occur (McArdle 1990), especially where they are rare, such as at the edge of their range.

Most current species distribution models (Guisan & Thuiller 2005) either require presence/absence data without attempting to separate true and false absences,

or they use presence-only data and do not use possible information on absences (but see Wintle *et al.* (2005) and Latimer *et al.* (2006) for static approaches incorporating detection probability). These models confound occurrence (Ψ) and the probability of detecting a species given that it occurs at a particular site (p). If $p < 1$, they underestimate Ψ. Furthermore, p probably varies spatially and can confound patterns in occurrence if not accounted for. Even if p is close to 1, static models (e.g. generalized linear models, GLMs) cannot directly relate the causal processes, such as extinction and colonization, to environmental variables. Our aim is to draw attention to occupancy models (MacKenzie *et al.* 2002, 2003, 2006) and their use for inferring species occurrence and its temporal dynamics in the face of imperfect detection.

2. MATERIAL AND METHODS

Occupancy models need repeated visits to a site and the information whether the species was recorded or not (1 or 0). A survey history is then constructed; for example, 101 represents the case where the species was detected on the first and third visit but not on the second visit. Evidently, the species occurs at this site, assuming the visits were made during a short time interval so that the species did not go locally extinct and recolonized the site during the study. The probability of observing the above survey history in terms of occupancy (Ψ) and detection rate (p_x for survey x) is $\Pr_{101} = \Psi p_1 (1 - p_2) p_3$. The survey history for a site where the species was never detected would be 000, and the probability of observing it is the sum of two possibilities; either the species was absent from the site or it was present but never detected: $\Pr_{000} = (1 - \Psi) + \Psi(1 - p_1)(1 - p_2)(1 - p_3)$. The likelihood of observing a set of survey histories is the product of the probabilities of observing each survey history, and the parameters (Ψ and p) can be found by maximum-likelihood or Bayesian methods (MacKenzie *et al.* 2002). The parameter Ψ is the proportion of sites that are occupied, corrected for the probability that the species has been missed.

For modelling species occurrence in relation to climate or other factors, Ψ and p are modelled as functions of covariates (x_i), with coefficients β to be estimated,

$$\text{logit}(\Psi) = \beta_0 + \sum \beta_i x_i.$$

This model is similar to GLMs used for bioclimatic modelling, except that it accounts for imperfect detection, which may vary in relation to observed covariates (e.g. observer skills, visibility, habitat structure). Non-parametric relationships between Ψ (or p) and climatic variables can be used in this type of model (Gimenez *et al.* 2006), in analogy to the generalized additive models also often used for bioclimatic modelling.

Dynamics in occupancy can be examined with multi-season occupancy models (MacKenzie *et al.* 2003). For each season, the design is as described above. Between seasons, however, colonization and local extinctions can take place. Consider a sampling design where sites were visited twice per year for 3 years. One possible survey history could be 01 00 10. The species was detected during the second visit of the first year and the first visit of the third year. It was not detected during the second year. The species was either present during the whole time period but not detected during the second year, or it went extinct between the first and the second year and recolonized the site between the second and third year. The probability of observing this particular survey history is

$$\Pr_{01\ 00\ 10} = \Psi_1 (1 - p_{1,1}) p_{1,2} \{ (1 - \varepsilon_1)(1 - p_{2,1})(1 - p_{2,2})(1 - \varepsilon_2)$$

$$+ \varepsilon_1 \gamma_2 \} p_{3,1} (1 - p_{3,2}),$$

where ε and γ are extinction and colonization probabilities, respectively. For p, the first subscript indicates year and the second indicates survey number. Subscripts of other parameters indicate year. Extinction and colonization rates can be examined as a function of covariates, as mentioned previously. Occupancy during the first season (Ψ_1) is estimated directly and changes in occupancy are derived from the estimated extinction and colonization rates. Variants of the model directly estimate occupancy in each season and its dependence on covariates (MacKenzie *et al.* 2006).

These models can be fitted with the free softwares MARK (used here; White & Burnham 1999, http://welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm) and PRESENCE (Hines 2006, http://www.mbr-pwrc.usgs.gov/software/presence.html).
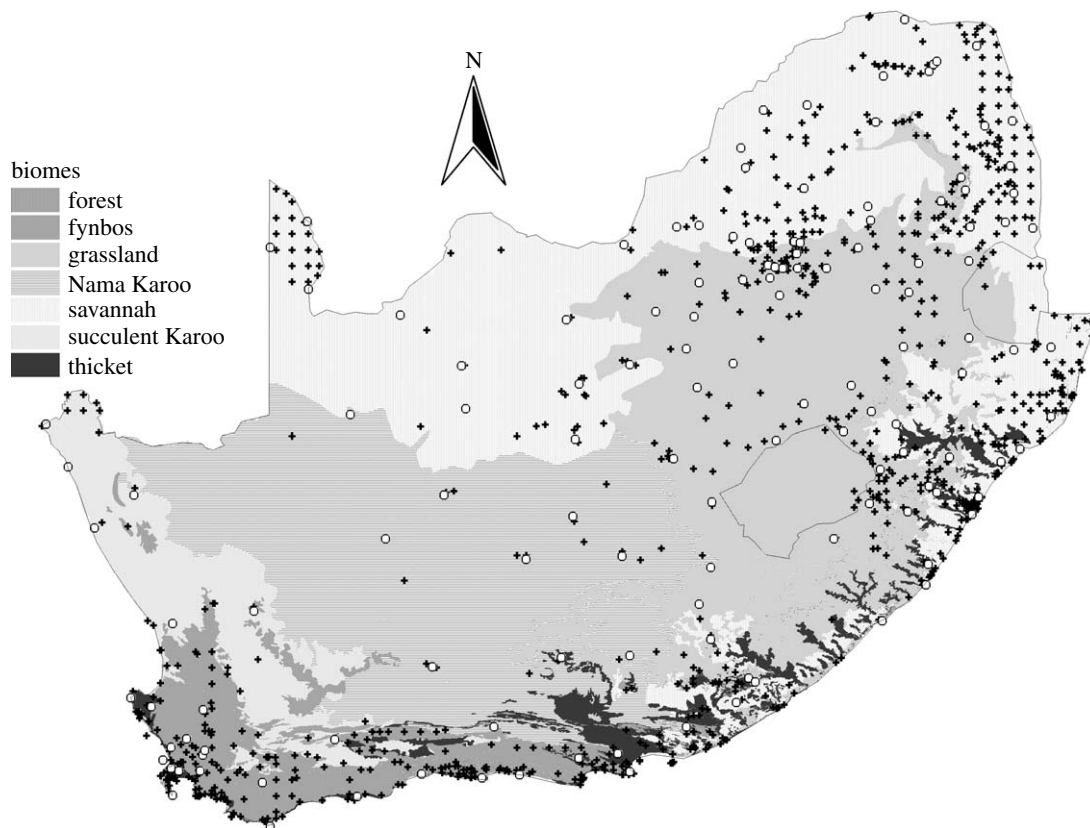
Figure 1. Map of South Africa/Lesotho/Swaziland showing the sites at which the data were collected (open circles, weather stations; plus symbols, Birds in Reserves Project (BIRP) sites) and the biomes used in the analysis. A further biome, not indicated in the keys, consists of near-shore islands.

(a) *Hadeda range dynamics*

We used occupancy models to examine range dynamics of the hadeda ibis (*Bostrychia hagedash*) in South Africa between 1994 and 2006. Hadedas have apparently expanded their range since 1900 (Macdonald *et al.* 1986). We used distribution data collected by the Birds in Reserves Project of the Animal Demography Unit, University of Cape Town (available at http://www.birds.sanbi.org/birp/birp_frameset_parent.htm). Volunteers visited 922 protected areas across South Africa between 1994 and 2006, and filled in a checklist of all the bird species they observed (figure 1). Not all reserves were visited in every year. There were 4526 reserve-by-year combinations with 1–153 checklists (median 3) collected each.

Hadedas feed mostly on invertebrates they extract from soft ground. As climatic variables affecting their distribution, we consider annual total rainfall, mean daily minimum temperature of the coldest month and mean daily maximum temperature of the hottest month, which we hypothesize affect access to food. We used mean climate as site-specific covariates on occupancy and annual deviations from the mean climate as year- and site-specific covariates on extinction and colonization. The weather data were collected at 136 weather stations across South Africa. We assigned each reserve to the nearest weather station (distance 0–149 km, median 26 km) using a geographical information system.

Furthermore, we expected all components of the model to vary between eight biome types recognized in South Africa: savannah (350 sites); grassland (215); forest (54); thicket (87); fynbos (144); Nama Karoo (20); succulent Karoo (28); and islands (24, treated separately here even though this is not a biome strictly speaking; hadedas rarely occur here).

3. RESULTS

Model selection favoured the model where initial occupancy depended on climate (model 3, table 1). The occupancy rate increased with higher minimum temperatures ($\beta = 0.010$, 95% confidence interval $(CI) = 0.006$–0.013), and tended to increase with decreasing maximum temperature ($\beta = -0.091$, $CI = -0.209$–0.027) and increasing rainfall ($\beta = 0.223$,

Table 1. Model selection results for multi-season occupancy analysis of the hadeda ibis (*Bostrychia hagedash*) in South Africa. (The model components are initial occupancy (Ψ), extinction (ε), colonization (γ) and detection probabilities (p). Subscripts indicate covariates (b, biome; y, year; clim, climate) in each model component. We based model selection on Akaike's information criterion (AICc; Burnham & Anderson 2002), where a smaller value indicates a better model (the best model in bold), and K is the number of parameters. Interactions are indicated by the multiplication sign, while the plus sign denotes models with main effects only.)

	model	AICc	K	deviance
1	$\Psi_b \varepsilon_b \gamma_b p_{b \times y}$	30 207.86	128	29 944.340
2	$\Psi_b \varepsilon_b \gamma_b p_{b+y}$	30 405.92	41	30 323.149
3	$\boldsymbol{\Psi_{clim+b} \varepsilon_b \gamma_b p_{b \times y}}$	**30 139.62**	**131**	**29 869.744**
4	$\Psi_{clim+b} \varepsilon_{b+y} \gamma_{b+y} p_{b \times y}$	30 161.41	153	29 844.621
5	$\Psi_{clim+b} \varepsilon_{clim+b} \gamma_{clim+b} p_{b \times y}$	30 141.12	137	29 858.490

$CI = -0.033$–0.479), as we expected. Adding annual variation in climate as a covariate on extinction and colonization did not improve the model (model 5, table 1).

Extinction and colonization varied among biomes, and the detection probability varied among biomes and years. However, we could not estimate all parameters for the Nama Karoo and islands, for which we had little data. In all other biomes, except forest and fynbos, colonization exceeded extinction, and hadedas thus occupied more sites over time. Occupancy declined in forests and fynbos (figure 2).
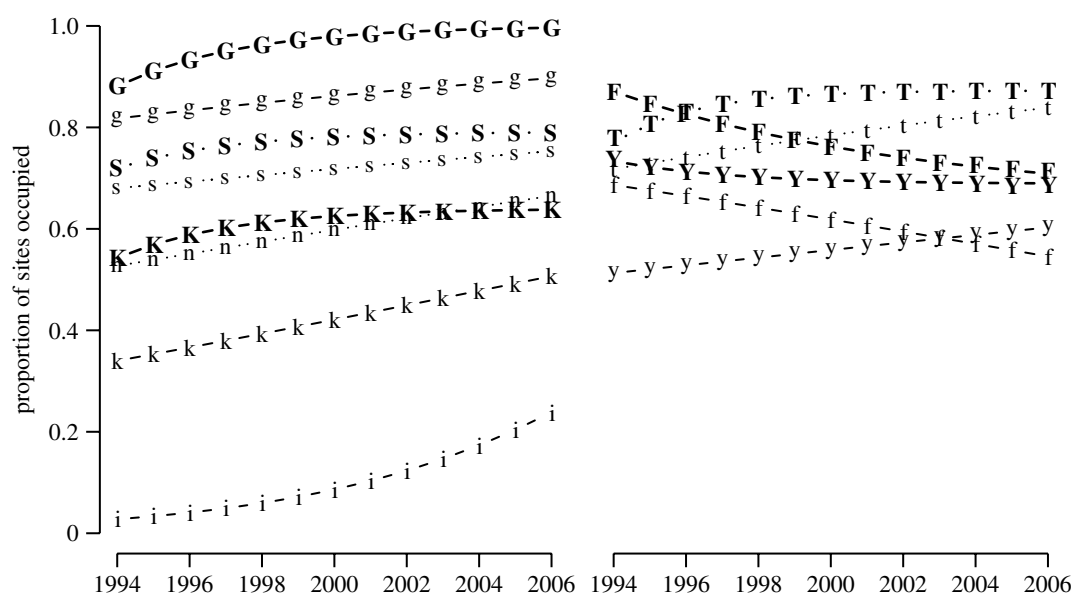
Figure 2. The proportion of South African protected areas occupied by hadedas between 1994 and 2006, predicted by occupancy model 3 given in table 1 (bold symbols) and a GLM with biome-specific time trends, but not accounting for detection probability (plain symbols). G,g, grassland; S,s, savannah; K,k, succulent Karoo; n, Nama Karoo; i, islands; T,t, thicket; F,f, forest; Y,y, fynbos.

We compared these results with a GLM, converting the checklists to presence/absence by years for each reserve. Not accounting for the detection probability, the GLM predicted lower occupancy rates, especially in the succulent Karoo and forest (figure 2). The GLM suggested an increased occupancy rate over time in fynbos, in contrast to the occupancy model, which suggests that the detection rate rather than occupancy increased over time. A time trend in detection rate could be due to increasing abundance in already occupied sites.

4. DISCUSSION

Do species change their ranges in response to climate change? If so, how fast? Do they colonize new areas that have become climatically suitable and go extinct from areas that became climatically unsuitable? Does the climatic niche shift over time (Broennimann *et al.* 2007)? These are pressing questions, and together with suitable datasets, multi-season occupancy models can be used to answer them because they can be used to examine factors affecting the dynamics of species occurrence. Multi-season occupancy models have only just begun to be applied to large spatial scales (MacKenzie *et al.* 2006, p. 201; Eraud *et al.* 2007).

Occupancy models require repeated visits to a sample of the sites about which inference is to be drawn. Given that relatively large spatial and temporal scales are of interest in climate change biology, this requirement may appear difficult to meet. However, we think that many existing datasets are amenable to the approach. In our example, we used checklists collected by volunteers, and such data are collected in many countries, for example, for atlas projects (birds: Greenwood 2007). Occupancy models have also been used with transect data, taking stations along transects as replicated observations (MacKenzie *et al.* (2006), p. 201 used North American Breeding Bird Survey data).

Occupancy models are being developed rapidly. If heterogeneity in detection probabilities is mainly caused by variation in abundance, these models can be used to estimate abundance (Royle & Nichols 2003) from repeated presence/absence surveys (or counts; Royle 2004). Other extensions estimate species richness from repeated survey data (Dorazio *et al.* 2006) and could be used to relate changes in biodiversity to climate or land-use change.

Our aim was to draw attention to the recently developed occupancy models, and their usefulness for understanding the range dynamics of species that are imperfectly detected. An advantage of occupancy models is that the detection process is incorporated into the model and factors thought to affect detection can be accounted for.

Broennimann, O., Treier, U. A., Müller-Schärer, H., Thuiller, W., Peterson, A. T. & Guisan, A. 2007 Evidence of climatic niche shift during biological invasion. *Ecol. Lett.* **10**, 701–709. (doi:10.1111/j.1461-0248.2007.01060.x)

Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information–theoretic approach*, 2nd edn. New York, NY: Springer.

Dorazio, R. M., Royle, J. A., Söderström, B. & Glimskär, A. 2006 Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* **87**, 842–854. (doi:10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2)

Eraud, C., Boutin, J. M., Roux, D. & Faivre, B. 2007 Spatial dynamics of an invasive bird species assessed using robust design occupancy analysis: the case of the Eurasian collared dove (*Streptopelia decaocto*) in France.

J. Biogeogr. **34**, 1077–1086. (doi:10.1111/j.1365-2699.2006.01673.x)

Gimenez, O., Covas, R., Brown, C. R., Anderson, M. D., Brown, M. B. & Lenormand, T. 2006 Nonparametric estimation of natural selection on a quantitative trait using mark–recapture data. *Evolution* **60**, 460–466.

Greenwood, J. J. D. 2007 Citizens, science and bird conservation. *J. Ornithol.* **148**, S77–S124. (doi:10.1007/s10336-007-0239-9)

Guisan, A. & Thuiller, W. 2005 Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009. (doi:10.1111/j.1461-0248.2005.00792.x)

Hines, J. E. 2006 *PRESENCE2—software to estimate patch occupancy and related parameters*. Reston, VA: USGS—PWRC.

Huntley, B., Collingham, Y. C., Green, R. E., Hilton, G. M., Rahbek, C. & Willis, S. G. 2006 Potential impacts of climatic change upon geographical distributions of birds. *Ibis* **148**, 8–28. (doi:10.1111/j.1474-919X.2006.00523.x)

Latimer, A. M., Wu, S. S., Gelfand, A. E. & Silander, J. A. 2006 Building statistical models to analyze species distributions. *Ecol. Appl.* **16**, 33–50. (doi:10.1890/04-0609)

Macdonald, I. A. W., Richardson, D. M. & Powrie, F. J. 1986 Range expansion of the hadeda ibis *Bostrychia hagedash* in southern Africa. *S. Afr. J. Zool.* **21**, 331–342.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S. R., Royle, J. A. & Langtimm, C. A. 2002 Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255.

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G. & Franklin, A. B. 2003 Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology* **84**, 2200–2207. (doi:10.1890/02-3090)

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L. & Hines, J. E. 2006 *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Amsterdam, The Netherlands: Academic Press.

McArdle, B. H. 1990 When are rare species not there? *Oikos* **57**, 276–277. (doi:10.2307/3565950)

Royle, J. A. 2004 N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**, 108–115. (doi:10.1111/j.0006-341X.2004.00142.x)

Royle, J. A. & Nichols, J. D. 2003 Estimating abundance from repeated presence–absence data or point counts. *Ecology* **84**, 777–790. (doi:10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2)

Thomas, C. D. *et al.* 2004 Extinction risk from climate change. *Nature* **427**, 145–148. (doi:10.1038/nature02121)

White, G. C. & Burnham, K. P. 1999 Program MARK: survival estimation from populations of marked animals. *Bird Study* **46**, S120–S139.

Wintle, B. A., Kavanagh, R. P., McCarthy, M. A. & Burgman, M. A. 2005 Estimating and dealing with detectability in occupancy surveys for forest owls and arboreal marsupials. *J. Wildl. Manage.* **69**, 905–917. (doi:10.2193/0022-541X(2005)069[0905:EADWDI]2.0.CO;2)