Advocating better habitat use and selection models in bird ecology

Facundo Xavier Palacio[1,2]

[1] Sección Ornitología, División Zoología Vertebrados, Facultad de Ciencias Naturales y Museo y Consejo Nacional de Investigaciones Científicas y Técnicas, Paseo del Bosque s/n, B1900FWA, La Plata, Buenos Aires, Argentina.

[2] Corresponding author: facundo_palacio@fcnym.unlp.edu.ar

ABSTRACT: Studies on habitat use and habitat selection represent a basic aspect of bird ecology, due to its importance in natural history, distribution, response to environmental changes, management and conservation. Basically, a statistical model that identifies environmental variables linked to a species presence is searched for. In this sense, there is a wide array of analytical methods that identify important explanatory variables within a model, with higher explanatory and predictive power than classical regression approaches. However, some of these powerful models are not widespread in ornithological studies, partly because of their complex theory, and in some cases, difficulties on their implementation and interpretation. Here, I describe generalized linear models and other five statistical models for the analysis of bird habitat use and selection outperforming classical approaches: generalized additive models, mixed effects models, occupancy models, binomial N-mixture models and decision trees (classification and regression trees, bagging, random forests and boosting). Each of these models has its benefits and drawbacks, but major advantages include dealing with non-normal distributions (presence-absence and abundance data typically found in habitat use and selection studies), heterogeneous variances, non-linear and complex relationships among variables, lack of statistical independence and imperfect detection. To aid ornithologists in making use of the methods described, a readable description of each method is provided, as well as a flowchart along with some recommendations to help them decide the most appropriate analysis. The use of these models in ornithological studies is encouraged, given their huge potential as statistical tools in bird ecology.

KEY-WORDS: binomial mixture models, classification trees, generalized additive models, generalized linear models, mixed models, occupancy models, regression trees.

HABITAT USE AND SELECTION IN BIRDS: FROM THEORY TO MODEL FIT

In the last decades, habitat use and selection has emerged as a basic aspect of bird ecology, due to its importance in natural history, distribution, response to environmental changes, management and conservation of bird species (Cody 1985, Guisan & Thuiller 2005, Engler *et al.* 2017). Despite the long tradition of the study of habitat use and selection in birds, however, almost 20 years ago, Jones (2001) had noticed ornithologists usually tended to be inconsistent of what habitat use and selection represent, with major implications on their hypothesis and conclusions about bird ecology (Jones 2001). Currently, some confusion between these terms still persists as a general issue in animal ecology (Lele *et al.* 2013, Boyce *et al.* 2016, McGarigal *et al.* 2016). Here, "habitat" is defined as a distinctive set of physical environmental factors that a species uses for survival and reproduction (Jones 2001, Lele *et al.* 2013). "Habitat use" refers to the way in which an individual or species uses habitats to meet its life history needs (Jones 2001).

"Habitat selection", by contrast, refers to a hierarchical process of behavioral responses that may result in the disproportionate use of habitats to influence survival and fitness of individuals (McGarigal *et al.* 2016). Therefore, habitat selection refers to a process, whereas habitat use refers to the pattern resulting from habitat selection (Jones 2001).

In the field, standard approaches to assess bird habitat use or selection involve: (1) sampling the presence or abundance of individuals of a species across sampling units (typically transects or point counts; Bibby *et al.* 2000) across different habitat types, (2) comparing presence locations with random locations where the species could potentially be present across different habitat types (use-availability or case-control approach; Jones 2001, Keating & Cherry 2004, Johnson *et al.* 2006), or (3) using tracking devices on individual birds to acquire location data and compare them to available locations where the species was not recorded (Burger & Shaffer 2008, Wakefield *et al.* 2009, Bridge *et al.* 2011). With the rise of powerful statistical methods and the advancement of computing facility, more complex designs have been developed

to assess habitat use and selection. For instance, these approaches can be extended to make repeated visits at the same sampling sites (temporal dependence), repeated observations on the same individuals (*e.g.* individuals tracked) or sampling many sites located nearby (spatial dependence). Notwithstanding, a plethora of statistical models outperforming classical linear models and which have been used for a while in other research areas (*e.g.* generalized linear and additive mixed models, Hastie & Tibshirani 1990, Bolker *et al.* 2009, Zuur *et al.* 2009; classification and regression trees, De'ath 2002, 2007; Ecological Niche Factor Analysis, Hirzel *et al.* 2002, Basille *et al.* 2008; quantile regression, Cade & Noon 2003; regularization methods such as ridge regression and LASSO, Reineking & Schröder 2006, James *et al.* 2013; Artificial Neural Networks, Lek & Guégan 1999; Flexible Discriminant Analysis, Hastie *et al.* 1994; Support Vector Machines, Kecman 2005; Bayesian approaches, Ellison 2004) are still not widespread among ornithologists. Some of these methods (*e.g.* generalized additive models, mixed models), nevertheless, have been widely used in some particular bird groups, such as seabirds (Wakefield *et al.* 2009, Engler *et al.* 2017). This phenomenon may be partly due to their relatively complex theory, and in some cases, difficulties on their implementation and interpretation (Bolker *et al.* 2009, Zuur *et al.* 2009, Dahlgren 2010). This is accentuated for Bayesian modeling, which represents a completely different statistical paradigm (Dennis 1996, Dorazio 2016). Moreover, early-career researchers tend to be reluctant to new analytical methods, as a result of self-perceived lack of quantitative training (Barraquand *et al.* 2014). Despite these issues, the methods mentioned typically both offer greater insight than classical approaches and represent no longer a problem in terms of statistical assumptions (Elith *et al.* 2006, Bolker *et al.* 2009, Elith & Graham 2009, Shabani *et al.* 2016).

From a statistical view, habitat use models aim to identify environmental variables linked to a species presence or abundance, and are species distribution models by definition (Guisan & Zimmerman 2000). On the other hand, habitat selection models link environmental variables with some proxy of fitness (nest site location, territories, reproductive output; Jones 2001). Although both types of models represent a correlative relationship between a bird species and its habitat, they are often expressed as a causal relationship, where the environment influences or explains the presence or abundance of a certain species:

$$\hat{Y} = b_0 + f(x)$$

where $\hat{Y}$ is the probability of occurrence or abundance of a bird species, b_0 is the intercept, x is an environmental variable, which may be represented by a categorical (different habitat types), ordinal or quantitative variable (*e.g.* environmental gradient), and $f(x)$ is a function of x. This simple model is suited for both habitat use and habitat selection studies, as it makes no assumptions of underlying processes, but just represents relationships between variables. It depends on the researcher whether this model is to be considered a habitat use or selection model (see Jones 2001). Beyond this theoretical discussion, the aim of this work is to describe some statistical methods appropriate for modeling the relationship between birds and their environment. As stated before, there is a myriad of methods that identify important environmental variables within a model, such as generalized additive models, mixed effects models, occupancy models, binomial mixture models and decision trees (classification and regression trees, bagging, random forests and boosting). In particular, these methods allow dealing with non-normal distributions (presence-absence and abundance data typically found in habitat use and selection studies), heterogeneous variances, non-linear relationships among variables, lack of statistical independence and imperfect detection. Here, I review these methods in order to (1) show the basics of each model with a readable description, (2) encourage ornithologists who are unfamiliar with the benefits of these methods to apply some of these analyses in their studies, and (3) help them to decide on which model to fit.

All graphs and models were built in R 3.3.1 (R Core Team 2016) using the packages lme4 (Bates *et al.* 2015), mgcv (Wood 2006), unmarked (Fiske & Chandler 2011), rpart (Therneau *et al.* 2015) and rpart.plot (Milborrow 2017).

REVIEW OF MODELING METHODS

Classical approaches: Generalized Linear Models

Generalized Linear Models (GLM) extend the classical linear regression approach by allowing different error distributions (not only normal) and the inclusion of non-homogeneous variances (Nelder & Wedderburn 1972). Every GLM has three basic components: (1) an error structure or random component, (2) a linear predictor or systematic component, and (3) a link function. The error structure corresponds to the distribution probability of the residuals (*i.e.* observed − predicted values), whereas the linear predictor represents the set of environmental variables. Finally, the link function $g\left(\hat{Y}\right)$ is a function of the response variable that links the error structure with the linear predictor, and makes the function linear (Dobson 2002):

$$g\left(\hat{Y}\right) = b_0 + b_1 x$$

where $\hat{Y}$ is the predicted occurrence or abundance of a bird species, $g(\hat{Y})$ is a function of $\hat{Y}$, x is an environmental variable, and b_0 and b_1 represent model coefficients. This simple model can be expanded to include non-linear effects through quadratic and interaction terms:

$$g(\hat{Y}) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2$$

where x_1 and x_2 represent environmental variables and b_i's represent model coefficients. Although GLMs can provide non-linear fits by including quadratic or cubic terms, they must use a high degree to produce flexible fits. Presence-absence data follow a binomial distribution (Pearce & Ferrier 2000), whereas count data may follow a Poisson, negative binomial or zero–inflated distributions (Welsh *et al.* 1996, Guisan *et al.* 2002, Ver Hoef *et al.* 2007). For the univariate case, binomial and Poisson GLMs are expressed, respectively, as:

$$\hat{Y} = \frac{\exp(b_0 + b_1 x)}{1 + \exp(b_0 + b_1 x)}$$

$$\hat{Y} = \exp(b_0 + b_1 x)$$

where b_0 and b_1 are model coefficients. Link functions transforming both models into a straight line are the logit $\log\left(\dfrac{\hat{Y}}{1-\hat{Y}}\right)$ and the log function, respectively (Dobson 2002):

$$\log\left(\frac{\hat{Y}}{1-\hat{Y}}\right) = b_0 + b_1 x$$

$$\log(\hat{Y}) = b_0 + b_1 x$$

The binomial GLM predicts the probability of occurrence of a species (Fig. 1), but the measured outcome is often codified as 0 (absent) and 1 (present). This model describes a logistic curve, and indeed, aroused independently from linear regression under the name of logistic regression (Hosmer-Jr. *et al.* 2013). The Poisson GLM, in contrast, predicts values between 0 and $+\infty$, as the response is represented by count data (Fig. 1). By using a Poisson distribution, heterogeneous variances are controlled, given that the expected value (mean) equals the variance. Thus, the larger the $\hat{Y}$, the larger the variance of the residuals (Fig. 1). However, in ecological data it is common for the variance to be larger than expected under a Poisson distribution (*e.g.* clumped distributions), which is termed "overdispersion" (Ver Hoef & Boveng 2007, Richards 2008, Lindén & Mäntyniemi 2011). Overdispersion may lead to wrong conclusions as

it inflates *P*-values, and thus it is imperative to control for it (Zuur *et al.* 2009). There are several ways to do so, which depends on the kind of data and amount of overdispersion. Essentially, it can be corrected by either including an overdispersion parameter (quasi-Poisson GLM) or using another distribution (negative binomial or zero-inflated; Potts & Elith 2006, Ver Hoef & Boveng 2007).

Examples of GLMs applied to birds include Oppel *et al.* (2012), who compared five modeling techniques, including GLMs, to predict the distribution of the Balearic Shearwater *Puffinus mauretanicus*. Rodríguez-Pastor *et al.* (2012) used a Poisson GLM to assess habitat use of the invasive Monk Parakeet *Myiopsitta monachus* in an urban area from Mexico, and Shahan *et al.* (2017) assessed the importance of local and landscape variables on grassland bird occurrence of prairie fragments using binomial GLMs.

Beyond linearity: Generalized Additive Models

GLMs establish relationships between the response and the environmental variables in a linear fashion. However, it is common for a species to show non-linear relationships with environmental variables, where species select environmental conditions in which they can survive and reproduce optimally. As a result, the presence or abundance of a species along an environmental gradient is usually unimodal (Austin 1987, Palmer & Dixon 1990). Generalized Additive Models (GAMs) extend GLMs by allowing the estimation of non-linear relationships between the response and the environmental variables, without assuming an *a priori* shape (Hastie & Tibshirani 1990, Yee & Mitchell 1991, Guisan *et al.* 2002). They are said to be data-driven instead of model-driven (like GLMs). A GAM is expressed as:

$$g(\hat{Y}) = b_0 + f(x)$$

where $f(x)$ is a non-linear function of x. Therefore, this model assumes no particular relationship between the response and the environmental variables. Like GLMs, GAMs can also use the same error distributions to model presence-absence and abundance data (binomial, Poisson, negative binomial, zero-inflated distributions) and link functions (logit, log; Yee & Mitchell 1991, Barry & Welsh 2002). Therefore, for presence-absence and count data, respectively, univariate GAMs are expressed as:

$$\hat{Y} = \frac{\exp[b_0 + f(x)]}{1 + \exp[b_0 + f(x)]}$$

$$\hat{Y} = \exp[b_0 + f(x)]$$

Basically, GAMs fit a smoothing curve by dividing the data into regions called "windows" at certain point locations called "knots", and then fit individual functions called splines within each window (Hastie & Tibshirani 1990, Zuur *et al.* 2009, James *et al.* 2013). There are many types of splines, but the most common involve cubic regression and smoothing splines. Cubic regression splines are cubic polynomials which are then joined together to form a smoothing curve (Zuur *et al.* 2009, James *et al.* 2013). Smoothing splines (also called penalized splines or *P*-splines) arise in a different situation, in which the aim is to find a function *f(x)* minimizing a residual sum of squares (*RSS*) subject to a smoothness penalty (Zuur *et al.* 2009, James *et al.* 2013):

$$\sum \left[Y - f\left(x\right) \right]^2 + \lambda J\left(f\right)$$

The first term is the *RSS* and measures the fit between the observed *Y* and expected values *f(x)*.

Minimizing only this term would lead to a function *f(x)* that exactly interpolates the data (James *et al.* 2013). Such a curve would be extremely rough, zigzagging among the different observed values, and with low predictive value. The second term is a penalty term measuring the degree of smoothness of the function *f(x)* (James *et al.* 2013), which is the product of a smoothness or tuning parameter λ and

$$J\left(f\right) = \int f''\left(x\right)^2 dx$$

where $f''(x)$ is the second derivative of *f(x)*. In practical terms, *J(f)* is a summed curvature of *f(x)*, which means it is a measure of roughness: it is large if *f(x)* is very wiggly, and it is zero if it is a straight line (James *et al.* 2013). If $\lambda = 0$, the penalty term has no effect, and *f(x)* will exactly interpolate the data. When λ is large, *f(x)* will be perfectly smooth (*i.e.* a straight line). Whatever type of spline is used, the result is that each environmental variable is
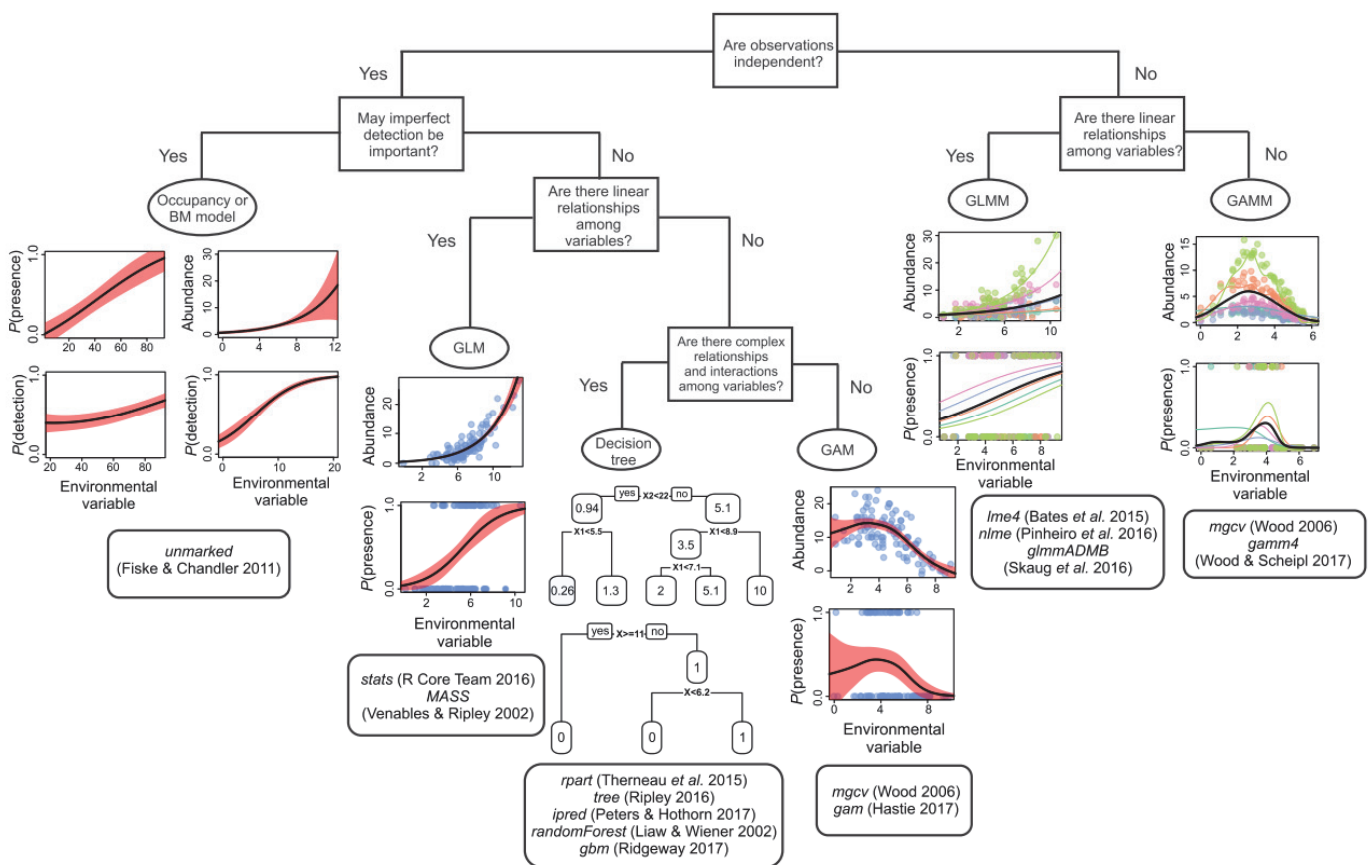
Figure 1. Flowchart illustrating how to select appropriate habitat use and selection models according to data properties. Rectangles hold questions, and ellipses contain recommended models. For most models (except for the classification tree), a hypothetical example using only one environmental variable is shown for the two most common data types in habitat models: presence-absence and abundance data. Points depict observations, black lines depict overall fitted models and red bands represent 95% confidence intervals. In decision trees, *Xs* represent environmental variables and values inside boxes, predicted values. Different colors in mixed models (GLMM and GAMM) indicate different levels of random effects (*e.g.* individuals, transects, point counts, etc.). Rounded corner boxes include R packages to perform the analysis. This is not a comprehensive review, but merely a guide to aid ornithologists to use an appropriate method. To choose a model, the researcher should also accompany this flowchart with data exploration, model validation and selection (see text for further details). BM - Binomial Mixture, GLM - Generalized Linear Model, GLMM - Generalized Linear Mixed Model, GAM - Generalized Additive Model, GAMM - Generalized Additive Mixed Model.

included in the model as a non-parametric smoothing function (Fig. 1). This only applies for numeric and ordinal data; nominal variables are included like in GLMs, in which case it is called a semi-parametric model, since it includes both non-parametric and parametric terms (James *et al.* 2013):

$$g\left(\hat{Y}\right) = b_0 + f(x_1) + b_1 x_2$$

where x_1 and x_2 are numeric and nominal, respectively, variables, and b_1 is a parametric coefficient.

GAMs, in contrast to GLMs, provide non-linear fits by increasing the number of knots but keeping the degree fixed (James *et al.* 2013). The main drawback of GAMs is that the fitted model is represented by a complex equation, and no coefficient estimates and standard errors are provided (Wood 2006, Zuur *et al.* 2009, Hegel *et al.* 2010). Instead, the significance of the model itself and the environmental variables is assessed, and a graphical display of the model relating environmental with response variables is often more useful (Wood 2006, Hegel *et al.* 2010).

Seabird ecology, contrary to terrestrial bird ecology, has often used GAMs as part of its statistical toolbox (Wakefield *et al.* 2009). As an example, Olivier & Wotherspoon (2006) assessed habitat selection in Wilson's Storm Petrel (*Oceanites oceanicus*) in both ice and ice-free areas where it nests. By using remote-sensing data, Fauchald *et al.* (2017) applied GAMs to relate foraging locations of Antarctic Petrels (*Thalassoica antarctica*) with melting ice and primary production. Pereira *et al.* (2018) used a combination of different models, including GAMs, to predict 30 seabird distributions as a function of different environmental stressors (fishing intensity, ship density and oil pollution risk). An example of the use of GAMs in terrestrial birds is Whitaker *et al.* (2015), who investigated the habitat use of the threatened Newfoundland Gray-cheeked Thrush (*Catharus minimus minimus*) at both local and landscape scales.

When assumptions of classical models are not met: mixed effects models

In habitat use and selection studies, it is often the case for ornithologists to take repeated samples of the same units (individual birds, point counts, transects, nests) over time (within a year, across years) or to have hierarchical or nested data (*e.g.* several samples of the same bird on different times). Under these scenarios of statistical non-independence, GLMs and GAMs are no longer valid, as these assume independence among observations (Hastie & Tibshirani 1990, Dobson 2002, Zuur *et al.* 2009). To cope with this, mixed effects models (or just mixed models; Bolker *et al.* 2009, Dingemanse

& Dochtermann 2013, Schielzeth & Nakagawa 2013) treat some factors grouping several observations that do not represent a directly measured effect (*e.g.* individual, point count, year, date, site, etc.) as random variables (*i.e.* random factors or effects). For the sake of simplicity, I will consider only one random factor (*e.g.* individuals or point counts with repeated observations). Fixed effects, on the contrary, represent the effects that explanatory variables have on the response variable and are supposed to be determined or fixed by the researcher (Bolker *et al.* 2009, Dingemanse & Dochtermann 2013), here represented by the environmental variables. Models are termed "mixed" because they include both random and fixed effects (Dingemanse & Dochtermann 2013) and are particularly valuable for identifying the source of unobserved variability and accounting for it, thus reducing the overall variance of the model (Bolker *et al.* 2009). Overall, mixed models extend GLMs and GAMs by including random effects, which are called, respectively, Generalized Linear Mixed Models (GLMMs) and Generalized Additive Mixed Models (GAMMs). As extensions of GLMs and GAMs, mixed models may use the same error and link functions.

In the simplest case, GLMMs and GAMMs can be expressed, respectively, as:

$$g\left(\hat{Y}\right) = b_0 + b_1 x + \gamma$$

$$g\left(\hat{Y}\right) = b_0 + f\left(x\right) + \gamma$$

Here, the intercept b_0 represents the grand mean of average individual or point count responses, whereas γ is each individual›s or point count›s unique average response (random effect) with γ coefficients normally distributed with a certain variance (Zuur *et al.* 2009, Dingemanse & Dochtermann 2013). In this model, the contribution of individuals or point counts is estimated as the difference from the population line by including intercepts for each individual or point count and keeping slopes constant (Dingemanse & Dochtermann 2013). This is called a random intercept model, as the intercepts of the individuals or point counts are assumed to be normally distributed with mean zero and variance σ^2 (Zuur *et al.* 2009, Dingemanse & Dochtermann 2013). This σ^2 represents the variance across random intercepts of individuals or point counts. In other words, in a random intercept GLMM there is an overall trend represented by the first two terms of the model (fixed effect), and one line fitted to each individual or point count parallel to the population fitted line (random effect), whose intercepts are assumed random (Fig. 1).

We may further suspect that the relationship between the environmental variables and the response is different for each individual or point count (*i.e.* they have different intercepts and slopes). By extension, intercepts

γ_1 and slopes γ_2 can vary randomly by assuming normal distributions with means zero and variances σ_1^2 and σ_2^2, respectively. This is called a random intercept and slope model (Zuur *et al.* 2009; Fig. 1):

$$g\left(\hat{Y}\right) = \left(b_0 + b_1 x\right) + \left(\gamma_1 + \gamma_2 z\right)$$

$$g\left(\hat{Y}\right) = \left[b_0 + f\left(x\right)\right] + \left(\gamma_1 + \gamma_2 z\right)$$

The first two terms represent the average relationship between presence or abundance and an environmental variable *x*, whereas the last two terms represent now individual curves for each of the point counts or individuals, whose intercepts γ_1 and slopes γ_2 vary randomly. In this model, *z* may be represented by different environmental variables including *x*, but this notation is used to distinguish the fixed component from the random one. The decision between both kinds of models is based on model selection or in biological meaning (see below; Zuur *et al.* 2009). It is worth mentioning that more complex designs exist, that allow including nested random factors, as well as spatial and temporal autocorrelation (nearby sampling units in space or time) commonly found in ornithological studies (Zuur & Ieno 2016). A comprehensive review is beyond the scope of this work. Under these scenarios, readers are encouraged to see Dormann *et al.* (2007), Zuur *et al.* (2009) and Zuur & Ieno (2016).

Mixed models have been used to fit the abundance and/or occurrence of birds to environmental variables. As examples, Paiva *et al.* (2010, 2013a, 2013b) analyzed foraging habitat use of different populations of Cory's Shearwater (*Calonectris diomedea borealis*) using Gaussian or binomial GLMMs with foraging trip nested within individual bird as random factor. Palacio (2016) assessed the habitat use of the Tufted-tit Spinetail (*Leptasthenura platensis*) in several forest remnants using a binomial GLMM with forest patch as random factor grouping occurrence records. Heldbjerg *et al.* (2017) analyzed GPS-location use of different land cover types as a function of the distance to the nest in the European Starling (*Sturnus vulgaris*) using a binomial GLMM with the identity of the individual as random factor. As in the case of GAMs, seabird ecologists have often used GAMMs to model habitat use and selection (Wakefield *et al.* 2009). Paiva *et al.* (2017), for instance, compared foraging habitat use derived from GPS-loggers between female and male Cory's Shearwaters in six breeding seasons. An example of GAMM applied to terrestrial birds is Sitters *et al.* (2014), who assessed the relationship between forest bird occurrence of 15 bird species, habitat structure and time since fire in a 70-year chronosequence using binomial GAMMs and specifying landscape (units of 100 ha separated by at least 3 km) and site (transects within landscapes) as random factors.

Accounting for imperfect detection: occupancy and binomial *N*-mixture models

So far, all the models described assume that detection of a species is perfect, that is, that every individual in the field is recorded. True absences are virtually impossible to assure, given the different sources of variation that may cause false negatives (a species may occur unnoticed either due to low abundance, cryptic or elusive habits, misidentification, or erroneous sampling designs; MacKenzie *et al.* 2005, Zuur *et al.* 2009, Guillera-Arroita 2017). Failure to account for imperfect detection may bias model parameter estimates, as the proportion of sites with presences (occupancy) will always underestimate the true occupancy, even with low levels of non-detection (Gu & Swihart 2003). Occupancy models account for imperfect detection by estimating both an occupancy probability ψ and a detection probability *p* based on making multiple visits at the same sample sites (called "detection histories"; MacKenzie *et al.* 2005, Guillera-Arroita 2017; Fig. 1). Mathematically, imperfect detection means $p < 1$. The essence of the method is that if a species is recorded in a given site (usually coded as 1), and then it is undetected (usually coded as 0), it is assumed that the site is occupied and the absences represent non-detections. As probabilities, both occupancy and detection are assumed to be binomially distributed (MacKenzie *et al.* 2005). In turn, each parameter may be constant or a function of one or more environmental variables, alike a binomial GLM (Welsh *et al.* 2013):

$$\psi = \frac{\exp\left(b_0 + b_1 x\right)}{1 + \exp\left(b_0 + b_1 x\right)}$$

$$p = \frac{\exp\left(\gamma_0 + \gamma_1 x\right)}{1 + \exp\left(\gamma_0 + \gamma_1 x\right)}$$

where ψ is occupancy probability, *p* is detection probability (given the species is present in a certain sampling site), *x* is an environmental variable (it may be the same or not for both occupancy and detection), and b_0, b_1, γ_0 and γ_1 are model coefficients (Fig. 1). This occupancy model may be one of the simplest approaches, and several extensions have been developed to expand this framework (*e.g.* multi-season: MacKenzie *et al.* 2003; multi-species: Dorazio *et al.* 2006, Richmond *et al.* 2010, Rota *et al.* 2016; alternative sampling designs: Lele *et al.* 2012, Specht *et al.* 2017).

A similar idea may be applied to deal with abundance instead of presence-absence data. Binomial *N*-mixture models (or *N*-mixture models) estimate both abundance *N* and detection probability *p* from abundance data

(Dodd-Jr. & Dorazio 2004, Kéry *et al.* 2005, Royle *et al.* 2005; Fig. 1). As before, *p* is assumed to be binomially distributed, whereas *N* is assumed to follow some distribution for count data (Poisson, negative binomial, zero inflated; Kéry 2008, Joseph *et al.* 2009). Assuming a Poisson distribution with mean λ, both λ and *p* can be modeled as functions of environmental variables:

$$N \sim Poisson(\lambda)$$

$$\lambda = \exp(b_0 + b_1 x)$$

$$p = \frac{\exp(\gamma_0 + \gamma_1 x)}{1 + \exp(\gamma_0 + \gamma_1 x)}$$

where *N* is the expected number of individuals, λ is the mean expected abundance, *p* is detection probability, *x* is an environmental variable, and b_0, b_1, γ_0 and γ_1 are model coefficients.

Models accounting for imperfect detection have several assumptions, the most important being the independence among sampling sites and the closed state of occupancy or demographic closure (Kéry *et al.* 2005, MacKenzie *et al.* 2005), which means that occupancy does not change at a site within the sampling period. Despite being promising tools, occupancy models suffer from several caveats. The estimating equations often have multiple solutions and the estimates are unstable when data are sparse, making accurate inference difficult (Welsh *et al.* 2013). Most importantly, when detection depends on abundance, model estimates are biased with similar magnitude to those biases obtained when ignoring non-detection (Welsh *et al.* 2013). As Welsh *et al.* (2013) has shown in a simulation study, occupancy modeling is not always applicable and should not be used indiscriminately to account for imperfect detection. In particular, sparse data (*i.e.* low number of occupied sites) results in extreme fits (0 or 1) for both detection and occupancy, because small changes in the data have large effects on the estimated parameters. Also, when detection is suspected to depend on abundance, occupancy models perform poorly (Welsh *et al.* 2013; but see Guillera-Arroita *et al.* 2014). Recently, however, Specht *et al.* (2017) proposed an alternative sampling design in which each of the sites is visited once, and sites where the species is encountered in the first survey are visited an additional number of times to better estimate detection probability. This conditional design expending a greater relative effort at occupied sites leads to improved parameter estimates (Specht *et al.* 2017).

In the last years, numerous applications of occupancy modeling and, to a lesser extent, *N*-mixture models have been applied to bird habitat use and selection. For instance, Parashuram *et al.* (2015) related Forest Thrush (*Turdus lherminieri*) abundance to forest structure using a binomial *N*-mixture model, and Suwanrat *et al.* (2015) applied a beta-binomial mixture model to estimate the abundance of the secretive Siamese Fireback (*Lophura diardi*) from camera trapping data in pristine and degraded forests. Glisson *et al.* (2017) and Huber *et al.* (2017) modeled the occupancy probability of the endangered wetland bird Yuma Ridgeway's Rail (*Rallus obsoletus yumanensis*) and the Wood Warbler (*Phylloscopus sibilatrix*), respectively, as a function of different environmental and anthropogenic disturbance variables.

Complex interactions among variables but simple decision rules: decision trees

Additional tools to identify important environmental variables in habitat use and selection models are decision trees (Breiman *et al.* 1984, De'ath & Fabricius 2000, De'ath 2002). Decision trees are non-parametric models that predict responses by recursively splitting the space of predictors (environmental variables) into a number of simple regions, giving as a result a dichotomous branching tree showing the hierarchy of importance of predictors as well as the nature of interactions between variables (Breiman *et al.* 1984, De'ath & Fabricius 2000, De'ath 2002). The tree is built by repeatedly splitting the data, defined by a simple rule based on a single explanatory variable (Fig. 1). At each split, the data is partitioned into two mutually exclusive groups, each being as homogeneous as possible. At each level of the tree, the mean of the observations in the region to which it belongs is used to make predictions (De'ath & Fabricius 2000). If the predicted response is presence-absence data, the tree is called classification tree; if the response is quantitative the tree is named regression tree (Breiman *et al.* 1984, Zuur *et al.* 2007; Fig. 1). A major advantage of decision trees is their simple and attractive graphical output (Fig. 1). As such, there is some terminology associated with trees, much alike phylogenetic trees. The root represents the top of the tree and initial split, and the terminal nodes are called leaves. The interpretation of the tree is made as follows: start at the root, and ask a sequence of questions about the environmental variables. The interior nodes are labeled with questions, and the edges or branches between them labeled by the answers (Fig. 1). Usually, each question refers to only a single variable, and has a yes/no answer. Arbitrarily, if the answer is "yes", we proceed to the left branch; otherwise, we proceed to the right branch (Fig. 1). The mean value of a group of observations is given as a prediction at the end of a branch.

A major issue to deal with is that decision trees tend to overfit, as a result of high complexity (*i.e.* many leaves). Overfitting leads to poor predictions, but using a small tree might also result in a poor fit (James *et al.* 2013).

A common approach to reduce overfitting is to grow a very complex tree and then apply an algorithm to prune it (Breiman *et al.* 1984). Intuitively, the aim is to find a sequence of subtrees of decreasing size, each of which is the best of all trees of its size, and then select the tree that gives better predictions. Pruning may be carried out through a cost-complexity approach. The cost can be defined by a metric such as the residual sum of squares (*RSS*) with a complexity penalty based on the size of the tree (Zuur *et al.* 2007):

$$RSS + cp \times tree\ size$$

The *RSS* measures data fit (see also GAMs), and *cp* is called complexity parameter ($cp \geq 0$). If the size of the tree is large, the *RSS* is relatively low and *vice versa* (recall that more complex trees tend to overfit thus reducing the *RSS*). The essence is to obtain a sequence of best subtrees as a function of *cp*, and then select the best subtree. *cp* values can be selected by cross-validation (Zuur *et al.* 2007, James *et al.* 2013), in which data are split in K (typically $K = 10$) subsets (K-fold cross validation). Each of these subsets is left out in turn, and a tree is computed for the remaining $(K - 1)/K$ percent of the data (90% if $K = 10$). Once the optimal tree size is calculated for a given *cp* value using the 90% subset, predictions are made and compared to observed values in the remaining 10% subset. The sum of squared differences between the observed and mean values per leaf is the prediction error. This process is applied for each of the $K = 10$ cross validations, giving 10 values of prediction errors. These 10 values are averaged for each *cp* value, and the *cp* value that minimizes the average prediction error is chosen (James *et al.* 2013).

Building a classification tree is quite similar to building a regression tree. However, *RSS* cannot be used in classification trees, as the outcome is 0 or 1. The Gini impurity index G can be used instead (Breiman *et al.* 1984):

$$G = \sum p\left(1 - p\right)$$

where p is the proportion of observations that belong to a given class. It is a measure of total variance across the K classes; if all observations belong to one class, $G = 0$ (no variance) and the node is considered "pure" as we can be pretty certain that the predicted values is either 0 or 1.

Main advantages of decision trees are: (1) their non-parametric nature, as they make no distributional assumptions about the data, (2) their simplicity and usefulness for interpretation, which make them ideal to explain to non-scientists (James *et al.* 2013), and (3) better at dealing with non-linearity and complex relationships between explanatory variables than other approaches

(GLM, GAM and mixed models; Zuur *et al.* 2007, James *et al.* 2013). On the other hand, they are less competitive in terms of statistical accuracy than other methods (*e.g.* GAMs). However, by aggregating many decision trees with methods like bagging, random forests, and boosting, the predictive performance of trees can be substantially improved (James *et al.* 2013).

A consequence of overfitting is that decision trees suffer from high variance. This means that splitting a dataset at random and fitting decision trees to each subset may give rather different results. To overcome this, bagging uses bootstrapping techniques to generate N different datasets (typically 100–1000), and then averages resulting predictions from each tree (Breiman 1996, De'ath 2007, James *et al.* 2013). Although bagging results in improved accuracy over prediction of a single tree, it is no longer possible to represent the results using one tree, and it is no longer clear which variables are the most important ones (James *et al.* 2013). However, it is possible to obtain a summary of the importance of each environmental variable by computing the decrease in *RSS* (regression trees) or Gini index (classification trees) due to splits over a given explanatory variable, averaged over all trees; a large value indicates an important predictor.

Now suppose the following setting: there is one very strong explanatory variable in the data set, and a number of other moderately strong predictors. In the set of bagged trees, most or all of the trees will use this strong predictor as first split. Consequently, all of the bagged trees will be similar to each other, and their predictions will be highly correlated. Averaging many highly correlated values does not substantially decrease variance as averaging many uncorrelated quantities. Another tree-based technique, called random forests, provides an improvement over bagging, as a way to reduce correlations between predictions of different trees (Breiman 2001, Cutler *et al.* 2007, James *et al.* 2013). As in bagging, a number of decision trees are built on bootstrapped samples, but each time a split in a tree is considered, a random sample of m predictors is chosen from the full set of p predictors. The number of m random predictors is typically fixed at $\sqrt{p}$.

A third approach for improving predictions of decision trees is boosting, also called boosted trees, in which each tree is grown sequentially using information from previous trees to improve error (De'ath 2007, Elith *et al.* 2008, James *et al.* 2013). Boosting assigns a weight to each model based on classification error. At each iteration, weights are increased on the incorrectly classified classes to focus the algorithm on these cases. The basic method proceeds as follows: given a current model, a decision tree is fitted using the residuals of the models as response. This new decision tree is added into the fitted function in order to update the residuals. Each of these trees can be small, with a few leaves. By fitting small trees

to the residuals, the function is slowly improved in areas where it does not perform well. Boosting does not involve bootstrapping; each tree is fitted on a modified version of the original dataset instead. For a detailed description of the method see De'ath (2007).

Decision trees applied to birds have been typically used in Ecological Niche Modeling (Engler *et al.* 2017). Examples include Marini *et al.* (2009, 2010), Quillfeldt *et al.* (2017), and Krüger *et al.* (2018), who used several models, including GAMs, classification trees, boosting and random forests to predict the abundance of the Red-spectacled Amazon (*Amazona pretrei*), 26 bird species from South America, the Black-browed Albatross (*Thalassarche melanophris*), and seven large seabird species of the Southern Ocean, respectively. Carrasco *et al.* (2017) used random forests to analyze the presence of breeding colonies in six species of herons and egrets as a function of land-use variables, and Steel *et al.* (2017) assessed habitat use in 15 terrestrial birds across a vineyard-matorral landscape using boosted classification trees.

HOW TO CHOOSE THE RIGHT MODEL?

After presenting some methods to analyze bird habitat use and selection, the obvious question is: "Which model should I fit to my data?" In an attempt to answer this question, I provide some general guidelines for ornithologists to decide on which model to use, partly summarized in Fig. 1. In the next sections, I will also briefly describe three broad issues in order to help researchers to recognize, at least, an appropriate model: (1) data types, sampling design and biological knowledge, (2) data exploration and model validation, and (3) model selection. Researchers must be aware that more than one model may be used to fit a particular dataset. Alternatively, and although I have tried to cover the most important types of data and designs in ornithological studies, none of the methods presented could fit a dataset well. Under these circumstances, researchers are encouraged to see also other methods mentioned in the current review.

Data types, sampling design and biological knowledge

The nature of the data gathered will undoubtedly have consequences on the type of model it can be applied, since it leads immediately to a subset of possible probability distributions to be handled. The two most common data types in habitat models are presence-absence and count data. As stated above, presence-absence data follow a binomial distribution, whereas count data may follow a Poisson, negative binomial or zero-inflated distributions. Zero-inflated Poisson (ZIP) or negative binomial (ZINB) distributions will be particularly useful for cryptic and

rare birds, in which there are an excess of zeroes and a low number of records (Welsh *et al.* 1996, Martin *et al.* 2005, Zuur *et al.* 2012). Extensions to account for imperfect detection have also been developed (Wenger & Freeman 2008, Joseph *et al.* 2009, Dénes *et al.* 2015). Although both data types are, by far, the most widely used in habitat models, they are not the only ones. In seabirds, for example, it is common to assess habitat use using proxies of foraging activity and distribution (*e.g.* trip duration, time required for a bird to pass through a circle with a given radius–first passage time duration–, foraging area, home ranges of foraging excursions; Pinaud 2007, Paiva *et al.* 2013b, 2017), which are continuous variables following Gaussian or beta distributions.

Sampling design and field methods are other major drivers of the model to be applied (Zuur & Ieno 2016). In bird habitat selection studies, it is common to have one or more sources of dependency. For example, transects or point counts are usually visited multiple times within a season, a year or between different years. These multiple visits represent a source of dependency, which may be modeled by mixed models (Zuur *et al.* 2009), occupancy or binomial *N*-mixture models. The same applies to repeated foraging observations of individual birds, in which the individual must be included as a random factor in a mixed model. Another valuable method used to model bird habitat selection is GPS tracking of individual birds, in which individual locations are dependent observations (Wakefield *et al.* 2009, Singh *et al.* 2016, Paiva *et al.* 2017). In this case, the individual bird is treated as a random factor in a mixed model.

Biological knowledge on the species under study has also implications for choosing a given model (Burnham *et al.* 2011). For instance, occupancy and binomial *N*-mixture models are ideal for terrestrial birds, for which there are many sources of variation impairing detection (see above), and, in particular, for cryptic or elusive species, for which $p << 1$ (Wenger & Freeman 2008). More rarely, occupancy models have been applied to seabirds. This is because the absence of vegetation and the size and conspicuousness of nesting colonies allows p being considered nearly or equal to 1 (Passuni *et al.* 2016). As an example, Passuni *et al.* (2016) assessed habitat selection with occupancy models in breeding colonies of three tropical seabird species and its relationship with oceanographic conditions and prey availability. In mixed models, for instance, if a bird appears in flocks (or any other groups, such as colonies, roosts or leks, namely areas where males aggregate to perform competitive displays for the females) then the flock should be treated as a random factor, since the presence and abundance of a species depends on the movement of other individuals in the group (Avilés & Bednekoff 2007, Xu *et al.* 2010, Végvári *et al.* 2016).

Data exploration and model validation

A fundamental step in data analysis is data exploration, as it provides insight into the data and their limitations, helps the researcher to identify appropriate models and allows checking model assumptions (Zuur *et al.* 2010). In this sense, graphical tools are advocated as the most important devices for data exploration, whereas certain statistical tests are warned against (Quinn & Keough 2002, Läärä 2009). Zuur *et al.* (2010) provide a protocol for data exploration covering important issues in exploratory data analysis, such as heterogeneity of variance, dependence among observations, zero inflation in GLMs and types of relationships between the response and explanatory variables. In particular, visualization of model residuals represents a key step to check whether a model meets its assumptions (*i.e.* model validation; Quinn & Keough 2002, Zuur *et al.* 2010, Zuur & Ieno 2016). To this end, a plot of residuals *vs.* fitted values, residuals *vs.* each environmental variable, and residuals *vs.* time or space coordinates, if relevant, must be made (Zuur *et al.* 2009, 2010). In all these plots, residual variation should be similar, showing no pattern. Although sometimes the researcher may think observations are *a priori* independent (which justifies the use of a GLM, GAM, occupancy models or binomial *N*-mixture models, Fig. 1), residuals may show some pattern. In these cases, a GLMM or GAMM should be a better choice (Fig. 1; Zuur *et al.* 2009, Zuur & Ieno 2016). Under temporal or spatial dependence, a GLMM or GAMM with temporal or spatial autocorrelation structure may be needed (Zuur *et al.* 2009). Finally, the choice between a linear and an additive model is based on the type of relationship between the response and the environmental variables (Fig. 1). If the relationship is linear or quadratic, consider using a GLM or GLMM; for more complex relationships consider applying a GAM or GAMM (Fig. 1; Zuur *et al.* 2009, Zuur & Ieno 2016). For complex relationships and interaction effects, decision trees are appropriate models. Although these assume no independence among observations, however, they are sensitive to autocorrelation effects (Segurado *et al.* 2006).

Model selection

A great body of literature has been devoted to the topic of model selection in ecology during the last decade (Burnham & Anderson 2004, Johnson & Omland 2004, Whittingham *et al.* 2006, Diniz-Filho *et al.* 2008, Burnham *et al.* 2011, Warren & Seifert 2011, Aho *et al.* 2014, Mac Nally *et al.* 2018). Once the researcher has identified an appropriate habitat model, he/she must choose one or several alternatives among a set of candidate models. To this, there are major two algorithms: (1) stepwise model selection (Whittingham *et al.* 2006) and (2) information-

theoretic approaches (IT approaches; Burnham & Anderson 2002). Stepwise selection *sensu lato* operates by successive addition or removal of significant or non-significant terms (forward selection or backward selection, respectively). Others operate by forward selection but also check the previous term to see if it can now be removed (stepwise selection *sensu stricto*; Whittingham *et al.* 2006). Stepwise selection is considered a poor procedure and is not recommended anymore, because it includes bias in parameter estimation, inconsistencies among model selection algorithms, the problem of multiple hypothesis testing, and an inappropriate focus or reliance on a single best model (Whittingham *et al.* 2006). Also, they are not able to compare non-nested models. For all these reasons, I will focus on IT approaches.

IT methods provide measures of the strength of evidence for a set of hypotheses (*i.e.* statistical models) given the data (Burnham & Anderson 2002). These are called "information-theoretic" because they are based on Kullback-Leibler (K-L) information (also called K-L distance or divergence). In essence, K-L information represents the information loss when model g_i is used to approximate reality f (process that generated observed data), or, in other words, the distance between g_i and f (Burnham *et al.* 2011). Thus, the idea is to select the model in the set of R models that minimizes K-L information loss (Burnham *et al.* 2011). Akaike (1973) found a simple expression describing the information loss when fitting a model, called Akaike's Information Criterion (AIC):

$$AIC = -2 \ln L + 2K$$

L is the likelihood, *i.e.* the probability of a model given the data, and K is the number of parameters in the model. Conceptually, the expression describes a trade-off between goodness-of-fit (first term with a negative effect on AIC) and complexity (second term with a positive effect on AIC). So, the higher the fit and the lower the number of parameters, the lower the AIC (*i.e.* principle of parsimony). In practice, AIC is computed for each of the R candidate models and the model with the smallest AIC value is selected as "best" (Burnham *et al.* 2011). Thus, it is the AIC differences (ΔAIC) that are important for ranking the models:

$$\Delta AIC_i = AIC_i - AIC_{min} \qquad \text{for } i = 1, 2, 3, \ldots, R.$$

where AIC_{min} is the minimum of the AIC values for the R models. Akaike weights w_i are a measure of strength of evidence and represent the probability of each model given the data and the R models under consideration (Burnham *et al.* 2011). In the literature, it is common to discard models with ΔAIC < 2. This arbitrary cutoff rule is now known to be poor, and models within the 2–7 range have support and should rarely be dismissed

(Arnold 2010, Burnham *et al.* 2011). After this procedure, nevertheless, there might be substantial model selection uncertainty, which is quantified by Akaike weights (*e.g.* the best model has probability 0.3). Under these circumstances, inferences should be based on all the models in the set of best models, which can be done with model averaging (Burnham & Anderson 2002). Grueber *et al.* (2011) suggest model averaging when w_i of the best model < 0.9. Basically, average coefficients result from a weighted average of the coefficients that appear in the best models, where the weights are represented by Akaike weights (Grueber *et al.* 2011). At this point, there are two approaches to compute these averages (Grueber *et al.* 2011): (1) full-model averaging or zero method, in which parameters not included in a model are set to zero and included when averaging the coefficient estimates, or (2) conditional-model averaging or natural average method, in which only those parameters included in a model are used for averaging (Burnham & Anderson 2002). The choice between both approaches depends on the aim of the study; Nakagawa & Freckleton (2011) recommend full-model averaging when the aim of the study is to determine those factors with the strongest effect on the response variable. On the other hand, when there is a particular factor of interest and it is possible that this factor may have a weak effect compared to other covariates, conditional-model averaging should be used to avoid shrinkage towards zero (Nakagawa & Freckleton 2011).

As AIC provides a relative measure of model fit, many different types of models may be compared. The comparisons are only valid for models fitted to the same response variable, so nested or non-nested models can be compared (Burnham & Anderson 2002, Grueber *et al.* 2011). Model 1 is said to be nested in model 2 if the parameters in model 1 are a subset of the parameters in model 2. For instance, a random intercept GLMM with one environmental variable is nested within another random intercept GLMM with the same environmental variable plus a quadratic term of the same variable. Another example is a random intercept GAMM and a random intercept and slope GAMM with the same environmental variable. Thus, different structures of random factors may be compared in mixed models (Zuur *et al.* 2009). It should be noted, however, that comparisons between mixed models is an active area of research, as there is no current consensus of how to handle random factors (Müller *et al.* 2013, Schielzeth & Nakagawa 2013, Rocha & Singer 2018, Sciandra & Plaia in press). Following the previous reasoning, GLMs are nested within GLMMs (GLMs with random factors), GAMs are nested within GLMMs (GAMs with parametric coefficients and random factors), and GAMMs encompass all these types of models (GLMs, GAMs and GLMMs), which means that all these are nested (Zuur *et al.* 2009). In contrast, a

GAM and a regression tree represent non-nested models, but they also can be compared using AIC. Thereby, IT approaches are a useful way to compare the different models presented in this review.

Overall, some basic principles guiding the use of AIC may be summarized: (1) AIC is a relative measure of model parsimony, so it only has meaning when comparing AIC values for different models; lower AIC indicates a more parsimonious model, relative to a model with a higher AIC (Burnham & Anderson 2002, Burnham *et al.* 2011), (2) nested, as well as non-nested, models can be compared (Burnham & Anderson 2002, Grueber *et al.* 2011), (3) too many models should not be compared, because a model with the lowest AIC, that is not the most appropriate model, might be found by chance; competing models should be based on biological meaning (Burnham *et al.* 2011), (4) it is possible to have multiple models performing similarly to each other, which may lead or not to model averaging (Grueber *et al.* 2011), (5) models with small sample sizes (as a rule of thumb, when n/K < 40) should be compared with the AIC corrected for small sample sizes (AIC_c; Hurvich & Tsai 1989), which penalizes stronger for the number of parameters in the model than AIC:

$$AIC_c = AIC + \frac{2K\,(K+1)}{n-K-1}$$

where n is sample size, and (6) the model identified as the "best" model may still have low explanatory or predictive power, so its adequacy needs to be addressed (Mac Nally *et al.* 2018); this can be achieved with different measures of explained variance (pseudo-R^2, R^2_{GLMM}; Nakagawa *et al.* 2017) coupled with cross-validation (Mac Nally *et al.* 2018).

Finally, AIC is not the only information criterion to determine the amount of information contained in a given model, but at present, it is by far the most widely used in ecology (Symonds & Moussalli 2011, Mac Nally *et al.* 2018). Other alternatives proposed include the Bayesian information criterion (BIC), the deviance information criterion (DIC), and the Watanabe-Akaike information criterion (WAIC), among others (see Box 1 in Grueber *et al.* 2011). For the criticisms of these indices, readers are encouraged to see Spiegelhalter *et al.* (2002) and Murtaugh (2009).

CONCLUDING REMARKS

This review presents powerful tools to model habitat use and habitat selection in ornithological studies. A comprehensive review of the methods available is beyond the scope of this work. Instead, this contribution is intended to give a broad overview of some of the most relevant approaches to analyze relationships between

birds and its environment, some of which still remain underused by ornithologists. Many other methods are available to model bird habitat use or selection, some of which require presence-only data such as Environmental Envelope Models (Hijmans & Graham 2006), Maximum Entropy (Elith *et al.* 2011, Merow *et al.* 2013) or Ecological Niche Factor Analysis (Hirzel *et al.* 2002, Basille *et al.* 2008). Other promising but more complex algorithms derived from machine-learning theory are Artificial Neural Networks (Lek & Guégan 1999, Yen *et al.* 2004) and Support Vector Machines (Guo *et al.* 2005, Kecman 2005). Hopefully, this work will attract ornithologists' interest in using some of the techniques presented, who will undoubtedly achieve a quality leap. Overall, the use of these models in ornithological studies is encouraged, given their huge potential as statistical tools in bird ecology.

ACKNOWLEDGEMENTS

REFERENCES

Aho K., Derryberry D. & Peterson T. 2014. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* 95: 631–636.

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle, p. 267–281. In: Petrov B.N. & Csaki F. (eds.). *Proceeding of the Second International Symposium on Information Theory*. Budapest: Akademiai Kiado.

Arnold T.W. 2010. Uninformative parameters and model selection using Akaike's Information Criterion. *Journal of Wildlife Management* 74: 1175–1178.

Austin M.P. 1987. Models for the analysis of species' response to environmental gradients. *Vegetatio* 69: 35–45.

Avilés J.M. & Bednekoff P.A. 2007. How do vigilance and feeding by Common Cranes *Grus grus* depend on age, habitat, and flock size? *Journal of Avian Biology* 36: 690–697.

Barraquand F., Ezard T.H.G., Jørgensen P.S., Zimmerman N., Chamberlain S., Salguero-Gomez R., Curran T.J. & Poisot T. 2014. Lack of quantitative training among early-career ecologists: a survey of the problem and potential solutions. *PeerJ* 2: e285.

Barry S.C. & Welsh A.H. 2002. Generalized Additive Modelling and zero inflated count data. *Ecological Modelling* 157: 179–188.

Basille M., Calenge C., Marboutin E., Andersen R. & Gaillard J.M. 2008. Assessing habitat selection using multivariate statistics: some refinements of the ecological-niche factor analysis. *Ecological Modelling* 211: 233–240.

Bates D., Maechler M., Bolker B. & Walker S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.

Bibby C.J., Burgess N.D., Hill D.A. & Mustoe S.H. 2000. *Bird census*

techniques. London: Academic Press.

Bolker B.M., Brooks M.E., Clark C.J., Geange S.W., Poulsen J.R., Stevens M.H.H. & White J.S.S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24: 127–135.

Boyce M.S., Johnson C.J., Merrill E.H., Nielsen S.E., Solberg E.J. & Moorter B. 2016. Can habitat selection predict abundance? *Journal of Animal Ecology* 85: 11–20.

Breiman L. 1996. Bagging predictors. *Machine Learning* 24: 123–140.

Breiman L. 2001. Random forests. *Machine Learning* 45: 5–32.

Breiman L., Friedman J.H., Stone C.J. & Olshen R.A. 1984. *Classification and regression trees*. New York: Routledge.

Bridge E.S., Thorup K., Bowlin M.S., Chilson P.B., Diehl R.H., Fléron R.W., Hartl P., Kays R., Kelly J.F., Robinson W.D. & Wikelski M. 2011. Technology on the move: recent and forthcoming innovations for tracking migratory birds. *BioScience* 61: 689–698.

Burger A.E. & Shaffer S.A. 2008. Application of tracking and data-logging technology in research and conservation of seabirds. *Auk* 125: 253–264.

Burnham K.P. & Anderson D.R. 2002. *Model selection and multimodel inference: a practical information-theoretic approach.* New York: Springer.

Burnham K.P. & Anderson D.R. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research* 33: 261–304.

Burnham K.P., Anderson D.R. & Huyvaert K.P. 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65: 23–35.

Cade B.S. & Noon B.R. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* 1: 412–420.

Carrasco L., Toquenaga Y. & Mashiko M. 2017. Balance between site fidelity and habitat preferences in colony site selection by herons and egrets. *Journal of Avian Biology* 48: 965–975.

Cody M.L. 1985. *Habitat selection in birds.* Orlando: Academic Press.

Cutler D.R., Edwards-Jr. T.C., Beard K.H., Cutler A., Hess K.T., Gibson J. & Lawler J.J. 2007. Random forests for classification in ecology. *Ecology* 88: 2783–2792.

Dahlgren J.P. 2010. Alternative regression methods are not considered in Murtaugh (2009) or by ecologists in general. *Ecology Letters* 13: E7–E9.

De'ath G. 2002. Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology* 83: 1105–1117.

De'ath G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88: 243–251.

De'ath G. & Fabricius K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81: 3178–3192.

Dénes F.V., Silveira L.F. & Beissinger S.R. 2015. Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution* 6: 543–556.

Dennis B. 1996. Discussion: should ecologists become Bayesians? *Ecological Applications* 6: 1095–1103.

Dingemanse N.J. & Dochtermann N.A. 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology* 82: 39–54.

Diniz-Filho J.A.F., Rangel T.F.L.V.B. & Bini L.M. 2008. Model selection and information theory in geographical ecology. *Global Ecology and Biogeography* 17: 479–488.

Dobson A.J. 2002. *An introduction to generalized linear models.* London: CRC Press.

Dodd-Jr. C.K. & Dorazio R.M. 2004. Using counts to simultaneously estimate abundance and detection probabilities in a salamander

community. *Herpetologica* 60: 468–478.

Dorazio R.M. 2016. Bayesian data analysis in population ecology: motivations, methods, and benefits. *Population Ecology* 58: 31–44.

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.

Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bolliger J., Carl G., Davies R.G., Hirzel A., Jetz W., Kissling W.D., Kühn I., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr F.M. & Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609–628.

Elith J. & Graham C.H. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.

Elith J., Graham C.H., Anderson R.P., Dudík M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L.G., Loiselle B.A., Manion G., Moritz C., Nakamura M., Nakazawa Y., Overton J.McC.M., Peterson A.T., Phillips S.J., Richardson K., Scachetti-Pereira R., Schapire R.E., Soberón J., Williams S., Wisz M.S. & Zimmermann N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.

Elith J., Leathwick J.R. & Hastie T. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802–813.

Elith J., Phillips S.J., Hastie T., Dudík M., Chee Y.E. & Yates C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57.

Ellison A.M. 2004. Bayesian inference in ecology. *Ecology Letters* 7: 509–520.

Engler J.O., Stiels D., Schidelko K., Strubbe D., Quillfeldt P. & Brambilla M. 2017. Avian SDMs: current state, challenges, and opportunities. *Journal of Avian Biology* 48: 1483–1504.

Fauchald P., Tarroux A., Tveraa T., Cherel Y., Ropert-Coudert Y., Kato A., Love O.P., Varpe O. & Descamps S. 2017. Spring phenology shapes the spatial foraging behavior of Antarctic petrels. *Marine Ecology Progress Series* 568: 203–215.

Fiske I.J. & Chandler R.B. 2011. Unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* 43: 1–23.

Glisson W.J., Conway C.J., Nadeau C.P. & Borgmann K.L. 2017. Habitat models to predict wetland bird occupancy influenced by scale, anthropogenic disturbance, and imperfect detection. *Ecosphere* 8: e01837.

Grueber C.E., Nakagawa S., Laws R.J. & Jamieson I.G. 2011. Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology* 24: 699–711.

Gu W. & Swihart R.K. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation* 116: 195–203.

Guillera-Arroita G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography* 40: 281–295.

Guillera-Arroita G., Lahoz-Monfort J.J., MacKenzie D.I., Wintle B.A. & McCarthy M.A. 2014. Ignoring imperfect detection in biological surveys is dangerous: a response to 'fitting and interpreting occupancy models'. *PLoS ONE* 9: e99571.

Guisan A., Edwards-Jr. T.C. & Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157: 89–100.

Guisan A. & Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993–1009.

Guisan A. & Zimmermann N.E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–186.

Guo Q., Kelly M. & Graham C.H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling* 182: 75–90.

Hastie T. 2017. *gam: Generalized Additive Models*. R package version 1.14-4. https://CRAN.R-project.org/package=gam (Access on 25 September 2017).

Hastie T., Tibshirani R. & Buja A. 1994. Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association* 89: 1255–1270.

Hastie T.J. & Tibshirani R.J. 1990. *Generalized additive models*. London: CRC Press.

Hegel T.M., Cushman S.A., Evans J. & Huettmann F. 2010. Current state of the art for statistical modelling of species distributions, p. 273–311. In: Cushman S.A. & Huettmann F. (eds.). *Spatial complexity, informatics, and wildlife conservation*. New York: Springer.

Heldbjerg H., Fox A.D., Thellesen P.V., Dalby L. & Sunde P. 2017. Common Starlings (*Sturnus vulgaris*) increasingly select for grazed areas with increasing distance-to-nest. *PLoS ONE* 12: e0182504.

Hijmans R.J. & Graham C.H. 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. *Global Change Biology* 12: 2272–2281.

Hirzel A.H., Hausser J., Chessel D. & Perrin N. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83: 2027–2036.

Hosmer-Jr. D.W., Lemeshow S. & Sturdivant R.X. 2013. *Applied logistic regression*. New York: John Wiley & Sons.

Huber N., Kéry M. & Pasinelli G. 2017. Occupancy dynamics of the Wood Warbler *Phylloscopus sibilatrix* assessed with habitat and remote sensing data. *Ibis* 159: 623–637.

Hurvich C.M. & Tsai C.L. 1989. Regression and time series model selection in small samples. *Biometrika* 76: 297–307.

James G., Witten D., Hastie T. & Tibshirani R. 2013. *An introduction to statistical learning*. New York: Springer.

Johnson C.J., Nielsen S.E., Merrill E.H., McDonald T.L. & Boyce M.S. 2006. Resource selection functions based on use–availability data: theoretical motivation and evaluation methods. *Journal of Wildlife Management* 70: 347–357.

Johnson J.B. & Omland K.S. 2004. Model selection in ecology and evolution. *Trends in Ecology & Evolution* 19: 101–108.

Jones J. 2001. Habitat selection studies in avian ecology: a critical review. *Auk* 118: 557–562.

Joseph L.N., Elkin C., Martin T.G. & Possingham H.P. 2009. Modeling abundance using *N*-mixture models: the importance of considering ecological mechanisms. *Ecological Applications* 19: 631–642.

Keating K.A. & Cherry S. 2004. Use and interpretation of logistic regression in habitat-selection studies. *Journal of Wildlife Management* 68: 774–789.

Kecman V. 2005. Support vector machines–an introduction, p. 1–47. In: Wang L. (ed.). *Support vector machines: theory and applications*. Berlin: Springer.

Kéry M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk* 125: 336–345.

Kéry M., Royle J.A. & Schmid H. 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* 15: 1450–1461.

Krüger L., Ramos J.A., Xavier J.C., Grémillet D., González-Solís J., Petry M.V., Phillips R.A., Wanless R.M. & Paiva V.H. 2018. Projected distributions of Southern Ocean albatrosses, petrels and fisheries as a consequence of climatic change. *Ecography* 41: 195–208.

Läärä E. 2009. Statistics: reasoning on uncertainty, and the insignificance of testing null. *Annales Zoologici Fennici* 46: 138–157.

Lek S. & Guégan J.F. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling* 120: 65–73.

Lele S.R., Merrill E.H., Keim J. & Boyce M.S. 2013. Selection, use,

choice and occupancy: clarifying concepts in resource selection studies. *Journal of Animal Ecology* 82: 1183–1191.

Lele S.R., Moreno M. & Bayne E. 2012. Dealing with detection error in site occupancy surveys: what can we do with a single survey? *Journal of Plant Ecology* 5: 22–31.

Liaw A. & Wiener M. 2002. Classification and regression by random forest. *R News* 2: 18–22.

Lindén A. & Mäntyniemi S. 2011. Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology* 92: 1414–1421.

Mac Nally R.M., Duncan R.P., Thomson J.R. & Yen J.D.L. 2018. Model selection using information criteria, but is the "best" model any good? *Journal of Applied Ecology* 55: 1441–1444.

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.

MacKenzie D.I., Nichols J.D., Royle J.A., Pollock K.H., Bailey L. & Hines J.E. 2005. *Occupancy estimation and modelling: inferring patterns and dynamics of species occurrence.* San Diego: Academic Press.

Marini M.Â., Barbet-Massin M., Lopes L.E. & Jiguet F. 2009. Predicted climate-driven bird distribution changes and forecasted conservation conflicts in a Neotropical Savanna. *Conservation Biology* 23: 1558–1567.

Marini M.Â., Barbet-Massin M., Martinez J., Prestes N.P. & Jiguet F. 2010. Applying ecological niche modelling to plan conservation actions for the Red-spectacled Amazon (*Amazona pretrei*). *Biological Conservation* 143: 102–112.

Martin T.G., Wintle B.A., Rhodes J.R., Kuhnert P.M., Field S.A., Low-Choy S.J., Tyre A.J. & Possingham H.P. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters* 8: 1235–1246.

McGarigal K., Wan H.Y., Zeller K.A., Timm B.C. & Cushman S.A. 2016. Multi-scale habitat selection modeling: a review and outlook. *Landscape Ecology* 31: 1161–1175.

Merow C., Smith M.J. & Silander-Jr. J.A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36: 1058–1069.

Milborrow S. 2017. *Plot rpart models: an enhanced version of 'plot. rpart'.* R package version 2.1.2. https://CRAN.R-project.org/package=rpart.plot (Access on 25 September 2017).

Müller S., Scealy J.L. & Welsh A.H. 2013. Model selection in linear mixed models. *Statistical Science* 28: 135–167.

Murtaugh P.A. 2009. Performance of several variable-selection methods applied to real ecological data. *Ecology Letters* 12: 1061–1068.

Nakagawa S. & Freckleton R.P. 2011. Model averaging, missing data and multiple imputation: a case study for behavioural ecology. *Behavioral Ecology and Sociobiology* 65: 103–116.

Nakagawa S., Johnson P.C.D. & Schielzeth H. 2017. The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface* 14: 20170213.

Nelder J.A. & Wedderburn R.W.M. 1972. Generalized linear models. *Journal of the Royal Statistical Society* 135: 370–384.

Olivier F. & Wotherspoon S.J. 2006. Distribution and abundance of Wilson's Storm Petrels *Oceanites oceanicus* at two locations in east Antarctica: testing habitat selection models. *Polar Biology* 29: 878–892.

Oppel S., Meirinho A., Ramírez I., Gardner B., O'Connell A.F., Miller P.I. & Louzao M. 2012. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* 156: 94–104.

Paiva V.H., Geraldes P., Marques V., Rodríguez R., Garthe S. & Ramos J.A. 2013a. Effects of environmental variability on different trophic levels of the North Atlantic food web. *Marine Ecology Progress Series* 477: 15–28.

Paiva V.H., Geraldes P., Ramírez I., Meirinho A., Garthe S. & Ramos J.A. 2010. Oceanographic characteristics of areas used by Cory's Shearwaters during short and long foraging trips in the North Atlantic. *Marine Biology* 157: 1385–1399.

Paiva V.H., Geraldes P., Ramírez I., Werner A.C., Garthe S. & Ramos J.A. 2013b. Overcoming difficult times: the behavioural resilience of a marine predator when facing environmental stochasticity. *Marine Ecology Progress Series* 486: 277–288.

Paiva V.H., Pereira J., Ceia F.R. & Ramos J.A. 2017. Environmentally driven sexual segregation in a marine top predator. *Scientific Reports* 7: 2590.

Palacio F.X. 2016. Tree cover, patch structure and patch isolation correlate with patterns of Tufted Tit-spinetail (*Leptasthenura platensis*) occurrence in Espinal Forest remnants from east-central Argentina. *Ornitología Neotropical* 27: 211–215.

Palmer M.W. & Dixon P.M. 1990. Small-scale environmental heterogeneity and the analysis of species distributions along gradients. *Journal of Vegetation Science* 1: 57–65.

Parashuram D., Oppel S., Fenton C., James G., Daley J., Gray G., Collar N.J. & Dolman P. M. 2015. The Forest Thrush *Turdus lherminieri* prefers mature mesic forest with dense canopy. *Bird Conservation International* 25: 503–513.

Passuni G., Barbraud C., Chaigneau A., Demarcq H., Ledesma J., Bertrand A., Castillo R., Perea A., Mori J., Viblanc V.A., Torres-Maita J. & Bertrand S. 2016. Seasonality in marine ecosystems: Peruvian seabirds, anchovy, and oceanographic conditions. *Ecology* 97: 182–193.

Pearce J. & Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133: 225–245.

Pereira J.M., Krüger L., Oliveira N., Meirinho A., Silva A., Ramos J.A. & Paiva V.H. 2018. Using a multi-model ensemble forecasting approach to identify key marine protected areas for seabirds in the Portuguese coast. *Ocean & Coastal Management* 153: 98–107.

Peters A. & Hothorn T. 2017. *ipred: improved predictors.* R package version 0.9-6. https://CRAN.R-project.org/package=ipred (Access on 27 September 2017).

Pinaud D. 2007. Quantifying search effort of moving animals at several spatial scales using first-passage time analysis: effect of the structure of environment and tracking systems. *Journal of Applied Ecology* 45: 91–99.

Pinheiro J., Bates D., DebRoy S. & Sarkar D. 2016. *nlme: linear and nonlinear mixed effects models.* R package version 3.1-128. http://CRAN.R-project.org/package=nlme (Access on 27 September 2017).

Potts J.M. & Elith J. 2006. Comparing species abundance models. *Ecological Modelling* 199: 153–163.

Quillfeldt P., Engler J.O., Silk J.R.D. & Phillips R.A. 2017. Influence of device accuracy and choice of algorithm for species distribution modelling of seabirds: a case study using black-browed albatrosses. *Journal of Avian Biology* 48: 1549–1555.

Quinn G.P. & Keough M.J. 2002. *Experimental design and data analysis for biologists.* Cambridge: Cambridge University Press.

R Core Team. 2016. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/ (Access on 25 September 2017).

Reineking B. & Schröder B. 2006. Constrain to perform: regularization of habitat models. *Ecological Modelling* 193: 675–690.

Richards S.A. 2008. Dealing with overdispersed count data in applied ecology. *Journal of Applied Ecology* 45: 218–227.

Richmond O.M.W., Hines J.E. & Beissinger S.R. 2010. Two-species occupancy models: a new parameterization applied to co-occurrence of secretive rails. *Ecological Applications* 20: 2036–2046.

Ridgeway G. 2017. *gbm: generalized boosted regression models.* R package

version 2.1.3. https://CRAN.R-project.org/package=gbm (Access on 26 September 2017).

Ripley B. 2016. *tree: classification and regression trees.* R package version 1.0-37. https://CRAN.R-project.org/package=tree (access on 27 September 2017).

Rocha F.M.M. & Singer J.M. 2018. Selection of terms in random coefficient regression models. *Journal of Applied Statistics* 45: 225–242.

Rodríguez-Pastor R., Senar J.C., Ortega A., Faus J., Uribe F. & Montalvo T. 2012. Distribution patterns of invasive Monk Parakeets (*Myiopsitta monachus*) in an urban habitat. *Animal Biodiversity and Conservation* 35: 107–117.

Rota C.T., Wikle C.K., Kays R.W., Forrester T.D., McShea W.J., Parsons A.W. & Millspaugh J.J. 2016. A two-species occupancy model accommodating simultaneous spatial and interspecific dependence. *Ecology* 97: 48–53.

Royle J.A., Nichols J.D. & Kéry M. 2005. Modelling occurrence and abundance of species when detection is imperfect. *Oikos* 110: 353–359.

Schielzeth H. & Nakagawa S. 2013. Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution* 4: 14–24.

Sciandra M. & Plaia A. in press. A graphical model selection tool for mixed models. *Communications in Statistics - Simulation and Computation*

Segurado P., Araújo M.B. & Kunin W.E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* 43: 433–444.

Shabani F., Kumar L. & Ahmadi M. 2016. A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and Evolution* 6: 5973–5986.

Shahan J.L., Goodwin B.J. & Rundquist B.C. 2017. Grassland songbird occurrence on remnant prairie patches is primarily determined by landscape characteristics. *Landscape Ecology* 32: 971–988.

Singh N.J., Moss E., Hipkiss T., Ecke F., Dettki H., Sandström P., Bloom P., Kidd J., Thomas S. & Hörnfeldt B. 2016. Habitat selection by adult Golden Eagles *Aquila chrysaetos* during the breeding season and implications for wind farm establishment. *Bird Study* 63: 233–240.

Sitters H., Christie F., Di Stefano J., Swan M., Collins P. & York A. 2014. Associations between occupancy and habitat structure can predict avian responses to disturbance: implications for conservation management. *Forest Ecology and Management* 331: 227–236.

Skaug H., Fournier D., Nielsen A., Magnusson A. & Bolker B. 2016. *glmmADMB: generalized linear mixed models using "AD Model Builder".* R package version 0.8.3.3/r287. https://R-Forge.R-project.org/projects/glmmadmb/ (Access on 27 September 2017).

Specht H.M., Reich H.T., Iannarilli F., Edwards M.R., Stapleton S.P., Weegman M.D., Johnson M.K., Yohannes B.J. & Arnold T.W. 2017. Occupancy surveys with conditional replicates: An alternative sampling design for rare species. *Methods in Ecology and Evolution* 8: 1725–1734.

Spiegelhalter D.J., Best N.G., Carlin B.P. & Linde A.V.D. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society* 64: 583–639.

Steel Z.L., Steel A.E., Williams J.N., Viers J.H., Marquet P.A. & Barbosa O. 2017. Patterns of bird diversity and habitat use in mixed vineyard-matorral landscapes of central Chile. *Ecological Indicators* 73: 345–357.

Suwanrat S., Ngoprasert D., Sutherland C., Suwanwaree P. & Savini T. 2015. Estimating density of secretive terrestrial birds (Siamese Fireback) in pristine and degraded forest using camera traps and distance sampling. *Global Ecology and Conservation* 3: 596–606.

Symonds M.R.E. & Moussalli A. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* 65: 13–21.

Therneau T., Atkinson B. & Ripley B. 2015. *rpart: recursive partitioning and regression trees.* R package version 4.1-10. https://CRAN.R-project.org/package=rpart (Access on 25 September 2017).

Végvári Z., Valkó O., Deák B., Török P., Konyhás S. & Tóthmérész B. 2016. Effects of land use and wildfires on the habitat selection of Great Bustard (*Otis tarda* L.) – implications for species conservation. *Land Degradation & Development* 27: 910–918.

Venables W.N. & Ripley B.D. 2002. *Modern applied statistics with S.* New York: Springer.

Ver Hoef J.M.. & Boveng P.L. 2007. Quasi-Poisson *vs.* negative binomial regression: How should we model overdispersed count data? *Ecology* 88: 2766–2772.

Wakefield E.D., Phillips R.A. & Matthiopoulos J. 2009. Quantifying habitat use and preferences of pelagic seabirds using individual movement data: a review. *Marine Ecology Progress Series* 391: 165–182.

Warren D.L. & Seifert S.N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21: 335–342.

Welsh A.H., Cunningham R.B., Donnelly C.F. & Lindenmayer D.B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. *Ecological Modelling* 88: 297–308.

Welsh A.H., Lindenmayer D.B. & Donnelly C.F. 2013. Fitting and interpreting occupancy models. *PLoS ONE* 8: e52015.

Wenger S.J. & Freeman M.C. 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology* 89: 2953–2959.

Whitaker D., Taylor P.D. & Warkentin I.G. 2015. Gray-cheeked Thrush (*Catharus minimus minimus*) distribution and habitat use in a montane forest landscape of western Newfoundland, Canada. *Avian Conservation and Ecology* 10: 4.

Whittingham M.J., Stephens P.A., Bradbury R.B. & Freckleton R.P. 2006. Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology* 75: 1182–1189.

Wood S.N. 2006. *Generalized additive models: an introduction with R.* New York: CRC Press.

Wood S.N. & Scheipl F. 2017. *Generalized additive mixed models using "mgcv" and "lme4".* R package version 0.2-5. https://CRAN.R-project.org/package=gamm4 (Access on 26 September 2017).

Xu Y., Yang N., Wang Y., Yue B.S. & Ran, J.H. 2010. Roosting behavior and roost selection by Buff-throated Partridges *Tetraophasis szechenyii* during the breeding season. *Zoological Studies* 49: 461–469.

Yee T.W. & Mitchell N.D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science* 2: 587–602.

Yen P.P.W., Huettmann F. & Cooke F. 2004. A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. *Ecological Modelling* 171: 395–413.

Zuur A.F. & Ieno E.N. 2016. A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution* 7: 636–645.

Zuur A.F., Ieno E.N. & Elphick C.S. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1: 3–14.

Zuur A.F., Ieno E.N. & Smith G.M. 2007. *Analyzing ecological data.* New York: Springer.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. 2009. *Mixed effects models and extensions in ecology with R.* New York: Springer.

Zuur A.F, Saveliev A.A. & Ieno E.N. 2012. *Zero inflated models and generalized linear mixed models with R.* Newburgh: Highland Statistics Limited.