MACROECOLOGICAL METHODS

Imperfect detection impacts the performance of species distribution models

José J. Lahoz-Monfort*, Gurutzeta Guillera-Arroita and Brendan A. Wintle

School of Botany, University of Melbourne, Parkville, Melbourne, Vic 3010, Australia

ABSTRACT

Aim Species often remain undetected at sites where they are present. However, the impact of imperfect detection on species distribution models (SDMs) is not fully appreciated. In this paper we evaluate the influence of imperfect detection on the calibration and discrimination capacity of SDMs. We compare the performance of three types of SDMs: (1) a technique based on presence–absence data, (2) a technique based on presence–background data, and (3) a technique based on detection/non-detection data that accounts for imperfect detection.

Innovation We use simulations to evaluate the impacts of imperfect detection in SDMs. This allows us to assess model performance with respect to the true objective of the models: the estimation of species distributions. We study a range of scenarios of occupancy and detection based on ecologically plausible environmental relationships and identify the circumstances in which imperfect detection affects model calibration and discrimination. We show that imperfect detection can substantially reduce the inferential and predictive accuracy of presence–absence and presence–background methods that do not account for detectability. While calibration is always affected, the influence on discrimination depends on the relationship of detectability and environmental variables.

Main conclusions The performance of a model should be assessed with respect to its objectives. Comparative studies that intend to assess the performance of an SDM by evaluating its ability to predict detections rather than presences fail to reveal the benefits of accounting for detectability. Disregarding imperfect detection can have severe consequences for SDM performance, and hence for the estimation of species distributions. To date, this issue has been largely ignored in the SDM literature. Simultaneously modelling occupancy and detection does not necessarily require a greater sampling effort, but rather that data are collected so that they are informative about detectability. We recommend that consideration of imperfect detection become standard practice for species distribution modelling.

Keywords
AUC, calibration, detectability, discrimination, GLM, hierarchical occupancy modelling, logistic regression, Maxent, presence–absence, presence–background.

Correspondence: José J. Lahoz-Monfort, School of Botany, University of Melbourne, Parkville, Melbourne, Vic 3010, Australia. E-mail: jose.lahoz@unimelb.edu.au

INTRODUCTION

The modelling of species distributions is an important tool in ecology, species management and biodiversity conservation (Peterson *et al.*, 2011). Knowledge about the distribution of species is vital for informing and prioritizing conservation action, including effective conservation planning (Guisan &

Thuiller, 2005). Efficiently planning for and managing the inevitable impacts of climate change on biodiversity will require predictions about the future distributions of species (e.g. Kearney *et al.*, 2010; Kujala *et al.*, 2013).

The most widely used species distribution models (SDMs) are based on statistical correlations between species occurrence data and spatial environmental data, and these models are the focus

of this paper. Over recent decades a suite of correlative SDM techniques have been proposed (for a comparison of some methods see Guisan & Zimmermann, 2000, and Elith *et al.*, 2006) with the most commonly implemented methods falling into two broad classes according to the type of observational data they require: (1) methods that require information on both where the species was observed to be present and where it was not observed ('presence–absence' methods) and (2) methods that require only presence records and environmental information about the studied landscape ('presence–background' methods; sometimes referred to as 'presence-only' methods).

A general problem of most species distribution modelling techniques is that they do not account for imperfect detection of species. However, species often remain undetected in surveys at occupied sites, and imperfect detection is recognized to be an important issue for the monitoring of species (Yoccoz *et al.*, 2001; Kéry, 2002; Martin *et al.*, 2005; Kéry & Schmidt, 2008). Unless a sufficiently large sampling effort is invested at surveyed locations, imperfect detection will result in the recording of false absences in presence–absence data sets and the omission of presences in presence–background data sets. This issue can lead to incorrect inference about a species' distribution, with misleading predictions that reflect where the species is more or less likely to be detected rather than where it is more or less likely to occur (Kéry, 2011; Monk, 2013). There is, however, one occupancy modelling technique that explicitly accounts for imperfect detection (MacKenzie *et al.*, 2002, 2006; Tyre *et al.*, 2003) by jointly modelling the processes describing where the species occurs and its detection at occupied sites. Hereafter we will refer to this technique as 'occupancy–detection' modelling to highlight its hierarchical structure which separates the state and observation processes (Royle & Dorazio, 2008). Uptake of this modelling technique has occurred primarily in the monitoring literature, and it has not been widely recognized as a useful tool for species distribution modelling. However, there has been recent work towards bridging this gap (Altwegg *et al.*, 2008; Kéry *et al.*, 2010, 2013; Rota *et al.*, 2011; Comte & Grenouillet, 2013).

In this paper we use simulations to carry out a thorough evaluation of the influence of imperfect detection on the inferential and predictive accuracy of SDMs, comparing the performance of presence–absence and presence–background methods that disregard imperfect detection with that of occupancy–detection models that explicitly model the detection process. Despite some discussion of imperfect detection in the context of SDMs (e.g. Kéry *et al.*, 2010; Monk, 2013; Yackulic *et al.*, 2013) and some previous work assessing the impact of non-detection on parameter estimation of logistic regression models (Tyre *et al.*, 2003; Gu & Swihart, 2004) and presence-only methods (Dorazio, 2012), this topic has not received sufficient attention to date. Our study presents a detailed investigation of the implications of detectability in terms of both model calibration and discrimination capacity, two fundamental aspects of SDM performance (Pearce & Ferrier, 2000; Jiménez-Valverde *et al.*, 2013). Calibration refers to the match between predicted probabilities and observed proportions, and hence to how accurately

the environmental relationships are estimated, while discrimination refers to the ability to classify binary instances correctly. A key feature of our study is that we evaluate these two performance aspects with respect to the actual objective of the models: the estimation of species distributions. In a recent paper, Rota *et al.* (2011) set out to answer the question of whether modelling imperfect detection improves the discrimination capacity of SDMs based on an analysis of bird survey data. However, they assessed the ability of the models to classify detections and hence did not look at the problem in terms of the true process of interest (i.e. the distribution of the species). Their conclusion, that accounting for imperfect detection provides 'surprisingly' little improvement in the discrimination capacity of SDMs, is dangerously misleading, and is indeed creating confusion in the literature (e.g. see the discussion in Conkin & Alisauskas, 2013). A similar conclusion is drawn in a recent comparison of the occupancy–detection model with a suite of SDM methods that disregard detectability, based on the same type of assessment (Comte & Grenouillet, 2013). Because the aim of SDMs is to model species distributions, it is critical to assess their performance in executing exactly this task, and not according to another criterion such as in Rota *et al.* (2011) and Comte & Grenouillet (2013). Assessing how well a given method can predict detections is not the same as assessing how well a model can describe or predict the spatial distribution of occupied sites. One of the aims of our paper is to clarify this crucial point, highlighting the true value of SDMs that account for imperfect detection.

In our study we compare the performance of three SDM methods: logistic regression, Maxent and occupancy–detection modelling. The logistic regression (a type of generalized linear model; McCullagh & Nelder, 1989) uses presence–absence records to model the distribution of the species as a function of a linear predictor via the logit link function, assuming that all absence records correspond to true absences. Maxent (Phillips & Dudík, 2008; Elith *et al.*, 2011) uses covariate information from sites at which the species has been observed and a large random sample of the environment (known as the 'background') to estimate the distribution of a species that is consistent with the data, avoiding unsupported additional assumptions (i.e. it estimates the distribution of maximum entropy). The occupancy–detection model (MacKenzie *et al.*, 2002) is an extension of the logistic regression that simultaneously models the occupancy and the detection processes. For these two processes to be separately identifiable, the model requires some information about the detection process, for instance the detection/non-detection history for a species in repeat separate surveys at sites (MacKenzie *et al.*, 2002) or data on detections over an interval of time (or distance) in a single visit (Garrard *et al.*, 2008; Guillera-Arroita *et al.*, 2011). We selected the logistic regression and Maxent as two well-known techniques that are representative of presence–absence and presence–background methods, but our general findings are applicable to other modelling techniques that assume perfect detection.

Simulation is the best tool for our assessment, for a number of reasons. First, since data are simulated, we have perfect

knowledge about the latent occupancy status of the sites (presence/absence) and thus we can assess the true performance of the models with respect to the actual objective (the modelling of species distributions), avoiding misleading results derived from assessing the performance to predict detections instead (Rota *et al.*, 2011; Comte & Grenouillet, 2013). Second, we can explore a range of scenarios of environmental relationships for species occupancy and detectability in a fully controlled way and hence identify when and how imperfect detection would represent a problem. Third, we can simulate and analyse a large number of data sets for each scenario and thus extract general conclusions about the performance of the models, avoiding potential spurious results derived from sampling variability. Finally, simulated data will not suffer from some of the important issues identified in the SDM literature (e.g. sampling bias, modelling fundamental versus realized niche; Araújo & Guisan, 2006).

In summary, the aim of this paper is to address three key questions: (1) Does imperfect detection affect the estimation of environmental relationships and thus SDM calibration, and if so, by how much and under what circumstances? (2) What are the implications for the ability of the model to discriminate between locations of true presence and absence? (3) How do the impacts of imperfect detection translate into geographical space (i.e. the estimated distribution)? By answering these questions, we demonstrate the effect of ignoring imperfect detection when fitting SDMs, which leads us to a discussion about the implications of imperfect detection in the context of its applications.

METHODS

We analysed simulated observation data sets under plausible scenarios of strong and weak environmental effects of occupancy, and different correlation structures between detection probability and occupancy. We ran 100 simulations per scenario. All simulations, including data generation and model fitting were undertaken in R (R Development Core Team, 2011). We illustrate our results in terms of how well models characterize the 'true' relationships underlying the simulated data and how well maps of predictions on the geographical space obtained from simulations over a real landscape match the true underlying distribution maps.

Hereafter we use ψ_i to denote the probability that site i is occupied by the species and p_i for the probability that the species is detected at site i during a survey given that it is present (conditional detectability). We use $p_i^\star$ to denote the probability that the species is detected at site i after K surveys given that it is present, i.e. $p_i^\star = 1 - (1 - p_i)^K$. Finally, we note that the unconditional probability that the species is detected at a site after K surveys is $d_i = \psi_i p_i^\star$. This last variable will play an important role when interpreting the results.

Data generation

We simulated two relationships between the environment and species occupancy probability, and four relationships between the environment and species detection probability, resulting in the simulation of eight combinations of environment, occupancy and detectability over a landscape of 10,000 sites. We first generated two independent (and thus uncorrelated) uniformly distributed covariates (C_A and C_B), standardized to have zero mean and unit standard deviation. We defined species occupancy (ψ) as driven by covariate C_A following a logistic regression with two strengths of a positive linear relationship (Fig. S1.1 in Appendix S1 in Supporting Information):

'steep': $\beta_0 = 1$ and $\beta_1 = 3$,

'gentle': $\beta_0 = 0$ and $\beta_1 = 1$,

where $\mathrm{logit}(\psi_i) = \beta_0 + \beta_1 C_{A_i}$. The above values correspond to average occupancy over the landscape of 0.6 and 0.5, respectively.

We defined species detection probability (p) either as a constant, as a function of the same covariate that affects occupancy (C_A) or as a function of a different independent covariate (C_B). We tried two relationships with C_A, increasing and decreasing, to represent scenarios in which occupancy and detectability are positively or negatively correlated. In particular, we used the following linear relationships (Fig. S1.1 in Appendix S1):

'constant': $\alpha_0 = 0$, $\alpha_1 = 0$ and $\alpha_2 = 0$ (i.e. $p_i = 0.5$ in all sites),

'positive correlation': $\alpha_0 = 0$, $\alpha_1 = 1$ and $\alpha_2 = 0$,

'negative correlation': $\alpha_0 = 0$, $\alpha_1 = -1$ and $\alpha_2 = 0$,

'independent' (i.e. different covariate): $\alpha_0 = 0$, $\alpha_1 = 0$ and $\alpha_2 = 1$

where $\mathrm{logit}(p_i) = \alpha_0 + \alpha_1 C_{A_i} + \alpha_2 C_{B_i}$. All these cases correspond to an average detectability of 0.5. Using these relatively simple environmental relationships allows us to extract general conclusions about the effects of imperfect detection on the inference and predictions obtained from the three classes of model: presence–background, presence–absence and occupancy–detectability.

We simulated the sampling of 400 sites at random, with two replicate detection/non-detection surveys per site ($K = 2$). We first determined whether each site was occupied by the species as the outcome of a Bernoulli trial with probability ψ_i. In occupied sites, we modelled detections as the outcome of K independent Bernoulli trials with probability p_i. We assumed no false positive detections, and hence that all surveys to empty sites lead to non-detection. For the logistic regression and Maxent analyses, we further processed the data set by collapsing the outcome of both surveys into a single record ('1' if the species was detected in any of the two visits and '0' otherwise, with the information of non-detections discarded for the Maxent analysis). By taking this approach, we ensured that the data fed to all modelling methods represented the same amount of sampling effort, which provided the most favourable comparison conditions for the methods that disregard imperfect detection. For Maxent, we provided the covariate information for all 10,000 sites as background information.

Model fitting

We fitted the logistic regression models using the R function *glm*, the occupancy–detection models using the function *occu* from the R-package 'unmarked' (Fiske & Chandler, 2011) and the Maxent models using the function *maxent* from the R-package 'dismo' (Hijmans *et al.*, 2013). In the logistic regression and occupancy–detection analyses we fitted a set of candidate models, starting with the simplest model without covariates, and selected the best fitting model in terms of the Akaike information criterion (AIC) (Burnham & Anderson, 2002). For the first three detectability scenarios ('constant', 'positive correlation' and 'negative correlation') we included linear and quadratic terms of C_A. In the occupancy–detection model we included the covariate terms in the detectability component as well, fitting all possible model combinations of these relationships in ψ and p (three combinations for the logistic model and nine for the occupancy–detection model). The Maxent model was fitted restricting the features to 'linear' and 'quadratic', as well as without imposing restrictions. For the fourth detectability scenario ('independent') we included linear terms of C_A and C_B, as well as their interaction (five combinations for the logistic model and 25 for the occupancy–detection model). In the Maxent analysis with restrictions the feature 'product' was also allowed. In all scenarios, the default configuration values were used in Maxent (Phillips & Dudík, 2008; Elith *et al.*, 2011). We used Maxent's 'logistic output', which represents an index of suitability rather than a probability (Elith *et al.*, 2006).

Model evaluation: calibration and discrimination

We assessed model calibration by plotting the estimated environmental relationships, as well as producing calibration plots to display actual proportions of simulated occupied sites against estimated probabilities (Pearce & Ferrier, 2000). We assessed discrimination capacity by computing the area under the receiver operating characteristic (ROC) curve (Hanley & McNeil, 1982; Pearce & Ferrier, 2000; Jiménez-Valverde, 2012). We evaluated these two performance aspects with respect to the prediction of presence/absence (the latent process), which is the objective of the models, as well as with respect to the prediction of detection/non-detection (the observable process). The occupancy–detection model provides separate estimates for occupancy (ψ) and detectability per survey at occupied sites (p). From these, we derived the estimate for the unconditional detection probability ($d = \psi p^*$). Both the logistic and Maxent models only estimate one process, which we interpreted for our evaluation both as an estimate of detection (d) and as an estimate of occupancy (ψ), since this is the correspondence made by these methods by assuming perfect detection ($p^* = 1$).

AUC values were calculated based on 1000 randomly selected validation sites (different from those used for training). Using a large number of sites ensured little uncertainty in the AUC estimation due to sampling, and hence allowed us to extract general observations about the performance of the models. We compared our results with the AUC values that are obtained when the scoring rule is based on the underlying probabilities of occupancy and detection used to generate the simulated binary outcomes (presences and detections). This is the AUC that a perfectly calibrated model would achieve, and thus we will refer to it as the 'perfect calibration AUC'. For a given landscape (i.e. frequency distribution of covariate values), the perfect calibration AUC depends on the shape of the environmental relationship, as this dictates the frequency of the probabilities that are used as the scoring rule for discrimination (see Appendix S2; Jiménez-Valverde *et al.*, 2013). Our scenarios with 'steep' transition in ψ have a high perfect calibration AUC for presence/absence (0.940) given that the probabilities that discriminate well (high and low) are prevalent in the landscape. As a contrast, the perfect calibration AUC for our scenarios with a 'gentle' transition is 0.753. A well-calibrated model (i.e. that estimates the relationship with covariates well) will have an AUC that is close to the reference AUC from the corresponding perfectly calibrated model (i.e. best possible discrimination capacity for a given environmental relationship). However, the opposite is not necessarily true: values close to that reference AUC can be achieved by models that are not well calibrated, as long as they give a good estimate of the relative suitability ranking of the sites (Pearce & Ferrier, 2000; Lobo *et al.*, 2008).

Projections on geographical space

SDMs estimate relationships with the covariates in environmental space and these can be directly translated into geographical space, following the distribution of covariate values over the landscape. We simulated a virtual species to provide a visual demonstration of the potential impact that disregarding imperfect detection has in geographical space. Instead of the uniformly distributed covariate values used in the previous sections, we selected a real landscape in eastern Spain and derived two environmental covariates with ArcMap10 (ESRI, Redlands, CA, USA), '*elevation*' and '*minimum distance to rivers*', to be used as covariates C_A and C_B, respectively, in the scenarios defined above (see Appendix S3 for details). Hence we computed the probability of occupancy of our virtual species as increasing with elevation, and detectability was either constant, increasing with elevation, decreasing with elevation or a function of distance to river irrespective of elevation, a set of combinations that represent ecologically plausible relationships. We then randomly sampled the distribution of our virtual species and followed the methods described in the previous sections to analyse the resulting survey data and evaluate the performance of the models. Finally, we projected the environmental relationships estimated by each model for each scenario back into geographical space using ArcMap10, and assessed the impact of imperfect detection by comparing visually the estimated and 'real' distribution maps. We also identified for each map the (approximately) 30% of pixels with the best estimated habitat (higher probability of occupancy or suitability index).
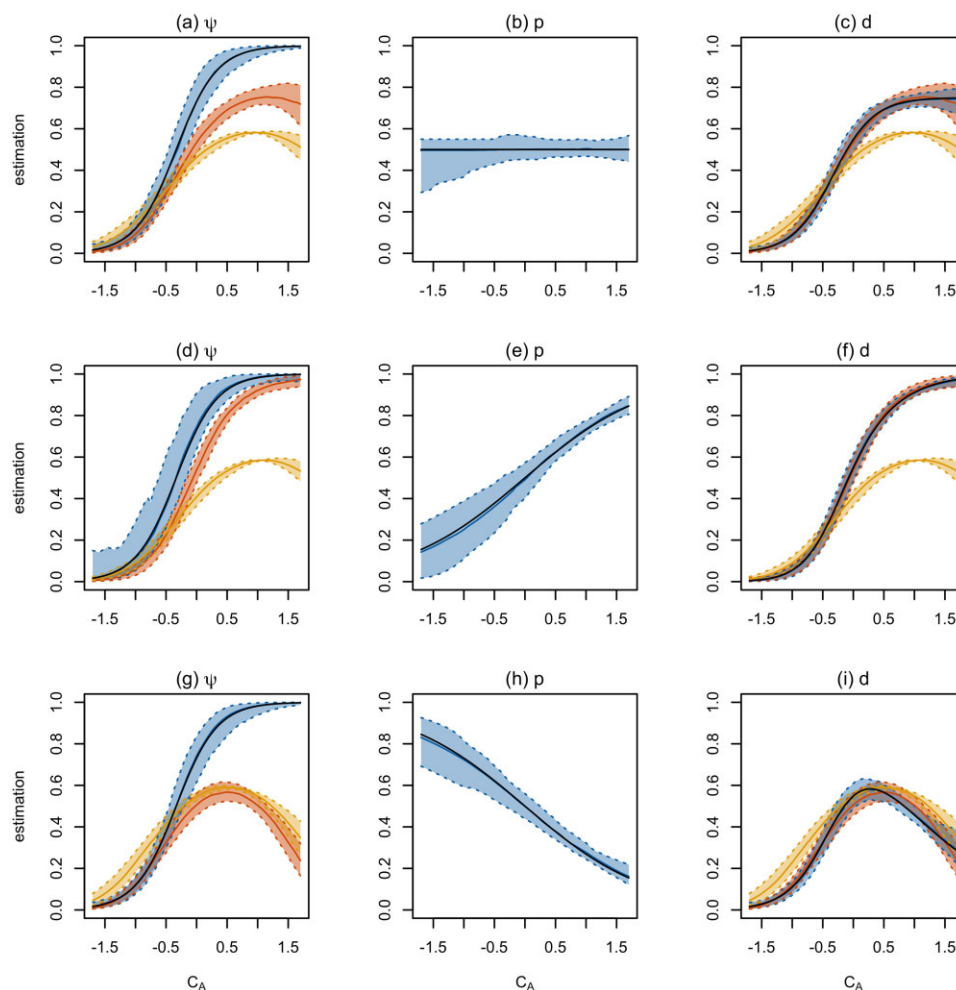
Figure 1 Environmental relationships with covariate C_A estimated with the occupancy–detection model (blue), the logistic regression model (red) and the Maxent model with linear and quadratic features (orange) for the 'steep' occupancy scenario and three detectability scenarios: 'constant' (top row), 'positive correlation' (mid row) and 'negative correlation' (bottom row); see parameter values in main text. The shaded areas represent the regions delimited by the 10th–90th percentiles of the estimates obtained from 100 simulations. From left to right, columns display the estimates of occupancy ψ, conditional detection probability in a survey p, and unconditional detectability d. Black lines show the true relationships. The estimates of the logistic and Maxent models are the same in the left and right columns. Maxent estimates represent a suitability index, not a probability.

RESULTS

Estimating environmental relationships (model calibration)

The results for the two occupancy relationships explored ('steep' and 'gentle') were similar. We report here the 'steep' scenarios (Figs 1 & 2); see Appendix S4 for the rest of the cases. Also, since as could be expected there were no significant differences in the performance of Maxent models with and without restrictions for the scenarios explored, we only display here the results for the former. Our simulations show that the logistic regression model gives a good estimate of the unconditional probability of species detection d (Fig. 1 column 3; Fig. 2c versus Fig. 2a2). However, if estimates are interpreted as occupancy probabilities, occupancy is underestimated and its relationship with the envi-

ronment is not well captured (Fig. 1 column 1; Fig. 2c versus Fig. 2a1). The occupancy–detection model on the other hand was able to tease apart true occupancy ψ from detection probability at occupied sites p, providing a good estimate of both processes and their relationships with the covariates, as well as of the derived unconditional probability of detection d (Fig. 1; Fig. 2b1, b2). Our plots clearly demonstrate that, unless imperfect detection is explicitly accounted for, rather than estimating where the species occurs we estimate where the species is detected. In other words, presence–absence models that do not explicitly incorporate detectability may be well calibrated to estimate detections but not so to estimate presence, which is the true variable of interest (see the calibration plots in Fig. 3). The estimates provided by the presence–background method, Maxent, do not represent probabilities but a suitability index and hence would not be expected to compare in magnitude with
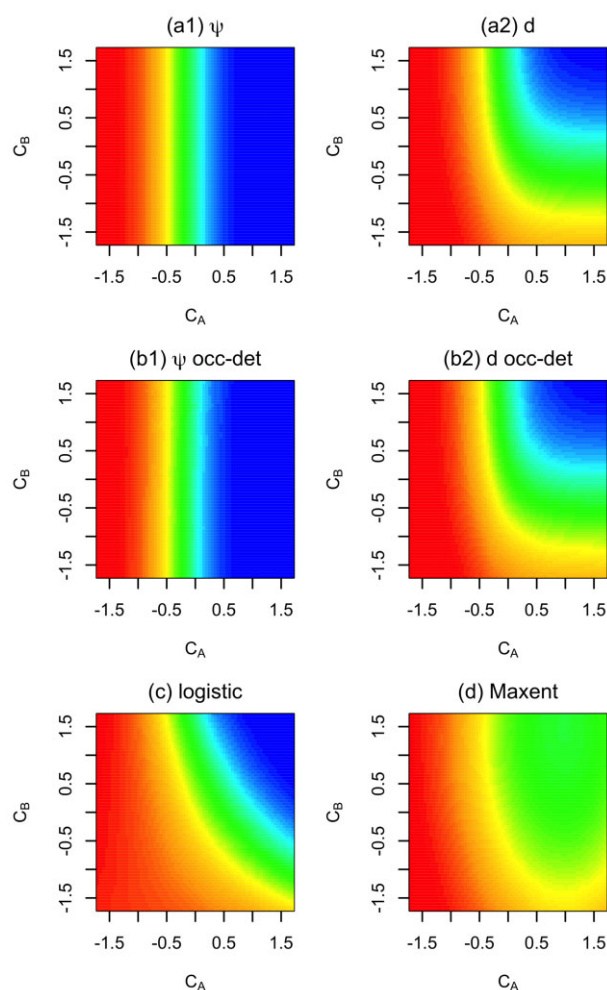
Figure 2 Environmental relationships with covariates C_A and C_B estimated with the occupancy–detection model (b1, b2, 'occ-det'), the logistic regression model (c, 'logistic') and the Maxent model with linear, quadratic and product features (d) for the 'steep' occupancy scenario when detectability and occupancy depend on two different covariates; see parameters values in main text. The true occupancy ψ and unconditional detection probabilities d are shown in (a1) and (a2), respectively. Colours range from red to dark blue corresponds to values from 0 to 1. Maxent estimates represent a suitability index, not a probability.

Figure 3 Calibration plots for the probability of occupancy ψ (left column) and the unconditional probability of detection d (right column) for the 'steep' occupancy scenario and four detection scenarios from top to bottom: 'constant', 'positive correlation', 'negative correlation' and 'independent'. Each colour represents one model: occupancy–detection model (blue), logistic regression model (red) and Maxent model with linear and quadratic features (orange). Each point represents the median of the observed proportions in the 100 simulations, and error bars indicate the 10th and 90th percentiles. Only the occupancy–detection model provides a well-calibrated output.

the true values of occupancy probability, even under perfect detection. Despite this, it can immediately be seen from the plots that Maxent predictions exhibit a very similar behaviour to those of logistic regression in terms of the shape of the estimated environmental relationship curves (Figs 1 & 2).

For the detectability scenarios 'constant' and 'positive correlation', the estimation of occupancy from the logistic and Maxent models increased monotonically (for most of its range; Fig. 1a,d). Hence the relative ranking of the occupancy estimates at the different sites was still correctly captured despite the underestimation and thus we would expect a discrimination performance similar to that of the occupancy–detection model. In contrast, in the 'negative correlation' scenario, both logistic
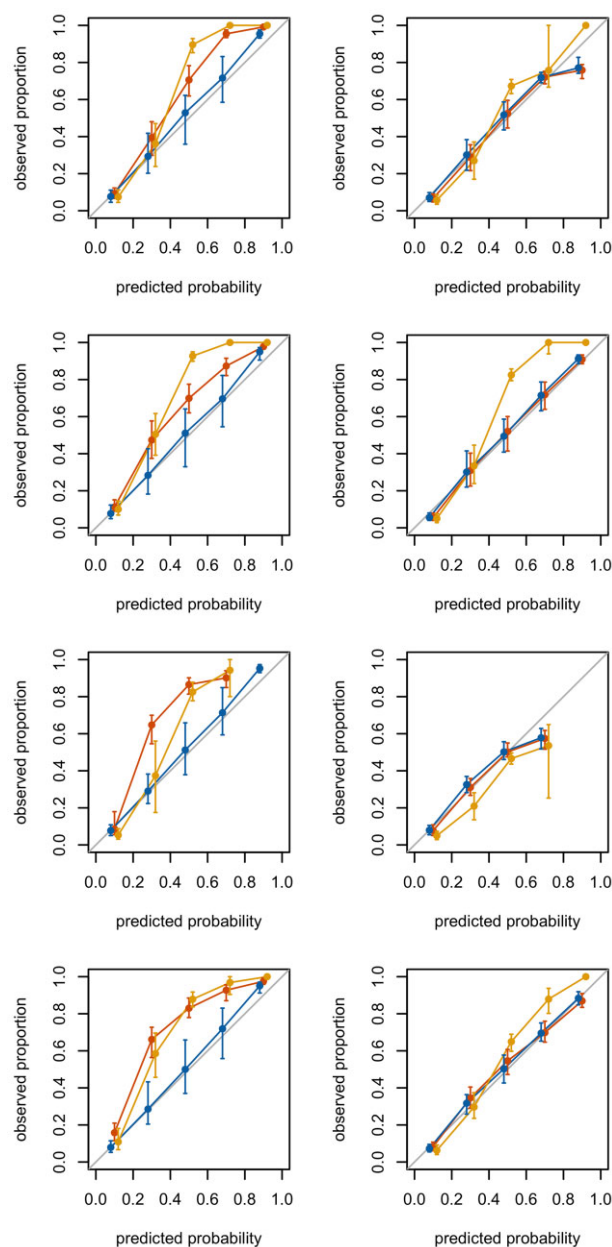
regression and Maxent models estimated a quadratic relationship with the covariate, resulting in a 'hump' in the curve, and thus the ranking of the sites was not maintained (Fig. 1g). In the 'constant' scenario, the selected logistic regression models also

Table 1 Median AUC (and 2.5th–97.5th percentiles) of the evaluated models for different scenarios of detectability p and occupancy ψ, obtained from 100 simulated data sets with 400 training sites, 1000 validation sites and two survey visits per site. Scenarios include two relationships of occupancy with a covariate C_A ('steep' and 'gentle') and four detectability scenarios (constant, positively correlated with ψ, negatively correlated with ψ and a function of a different covariate C_B); see parameter values in main text. AUC_{pres} refers to the capacity for discriminating true presence and absence, while AUC_{det} refers to the discrimination of detection and non-detection. The column 'perfect' shows the AUC corresponding to a perfectly calibrated model. The occupancy–detection model is superior for discriminating sites according to the true process of interest when detectability is either independent from or negatively correlated with occupancy. The lower 2.5th percentiles for the occupancy–detection model in the positive correlation cases are caused by a few of the simulations (5) in which the selected model was $\psi(.)p(C_A)$, i.e. all variability with the environment was interpreted by the model as caused by variation in detection.

Scenario		Presence/absence discrimination (AUC_{pres})					Detection/non-detection discrimination (AUC_{det})				
p	ψ	Perfect	Occ-det	Logistic	MaxentF	MaxentR	Perfect	Occ-det	Logistic	MaxentF	MaxentR
Constant	Steep	0.940	0.941 (0.923–0.953)	0.939 (0.924–0.953)	0.934 (0.900–0.951)	0.934 (0.915–0.950)	0.822	0.824 (0.801–0.849)	0.823 (0.801–0.850)	0.819 (0.791–0.845)	0.820 (0.796–0.847)
	Gentle	0.753	0.752 (0.712–0.776)	0.754 (0.731–0.776)	0.746 (0.716–0.773)	0.752 (0.725–0.776)	0.702	0.704 (0.677–0.740)	0.705 (0.677–0.740)	0.700 (0.671–0.734)	0.704 (0.672–0.738)
Covariate: positive correlation with ψ	Steep	0.940	0.941 (0.500–0.953)	0.941 (0.919–0.952)	0.939 (0.919–0.952)	0.939 (0.920–0.953)	0.921	0.922 (0.903–0.935)	0.922 (0.903–0.935)	0.916 (0.888–0.930)	0.916 (0.889–0.930)
	Gentle	0.753	0.752 (0.440–0.779)	0.754 (0.731–0.778)	0.751 (0.723–0.776)	0.753 (0.729–0.778)	0.807	0.808 (0.786–0.835)	0.808 (0.786–0.358)	0.805 (0.778–0.830)	0.806 (0.783–0.834)
Covariate: negative correlation with ψ	Steep	0.940	0.941 (0.908–0.953)	0.877 (0.819–0.923)	0.869 (0.790–0.917)	0.879 (0.815–0.929)	0.744	0.743 (0.714–0.775)	0.742 (0.712–0.773)	0.738 (0.690–0.775)	0.741 (0.706–0.773)
	Gentle	0.753	0.752 (0.729–0.779)	0.623 (0.493–0.768)	0.613 (0.463–0.744)	0.628 (0.461–0.756)	0.592	0.589 (0.540–0.628)	0.586 (0.500–0.628)	0.581 (0.529–0.626)	0.590 (0.546–0.628)
Independent from ψ	Steep	0.940	0.939 (0.911–0.951)	0.908 (0.871–0.930)	0.900 (0.852–0.929)	0.910 (0.871–0.934)	0.866	0.866 (0.846–0.883)	0.861 (0.838–0.878)	0.852 (0.825–0.877)	0.859 (0.836–0.878)
	Gentle	0.753	0.751 (0.709–0.776)	0.712 (0.654–0.748)	0.690 (0.622–0.731)	0.699 (0.633–0.743)	0.743	0.743 (0.710–0.771)	0.740 (0.705–0.767)	0.728 (0.690–0.760)	0.738 (0.704–0.766)

MaxentF, full set of features allowed; MaxentR, features restricted to 'linear' and 'quadratic', plus 'product' for the last scenario.

included a quadratic term, which was needed to capture the flattening of the unconditional probability of detection d at values smaller than 1.

For the scenario in which p and ψ depended on two independent covariates, the unconditional probability of detection (d) depended on both variables and therefore both logistic regression and Maxent favoured models with both covariates and their interactions (Fig. 2c, d). Unlike the single covariate scenarios, here the estimate of d in the logistic model differed substantially from the true values. This is because it is not possible to fully capture the structure of the true relationship with a single logistic equation.

Discrimination capacity

When detectability was constant, the ability to discriminate presence/absence (Table 1) was similar for all models, irrespective of their calibration. This was also the case when p was positively correlated with ψ. In contrast, there were considerable differences in AUC between models in the scenario of negative correlation between p and ψ, with the occupancy–detection model always achieving values close to those expected for a perfectly calibrated model and noticeably larger than those for models that disregard detectability (e.g. 0.94 vs. 0.87 for the 'steep' scenario; 0.75 vs. 0.62 for the 'gentle' scenario). The same effect was found when detectability varied independently from occupancy, although the differences in AUC were smaller (0.94 vs. 0.91 and 0.75 vs. 0.70; but note that these are true differences as in our study the AUC variances are negligible). We will see in the following section an illustration of the relevance of these AUC differences in practice.

The ability to discriminate detection/non-detection was equally good for all models. The AUC values were all similar and very close to those expected for a perfectly calibrated estimation of d (note that the 'perfect' AUC values now also depend on the environmental relationship with detectability p). We can conclude that all the models have good discrimination capacity for the detections, but again we note that this is not the true process of interest.
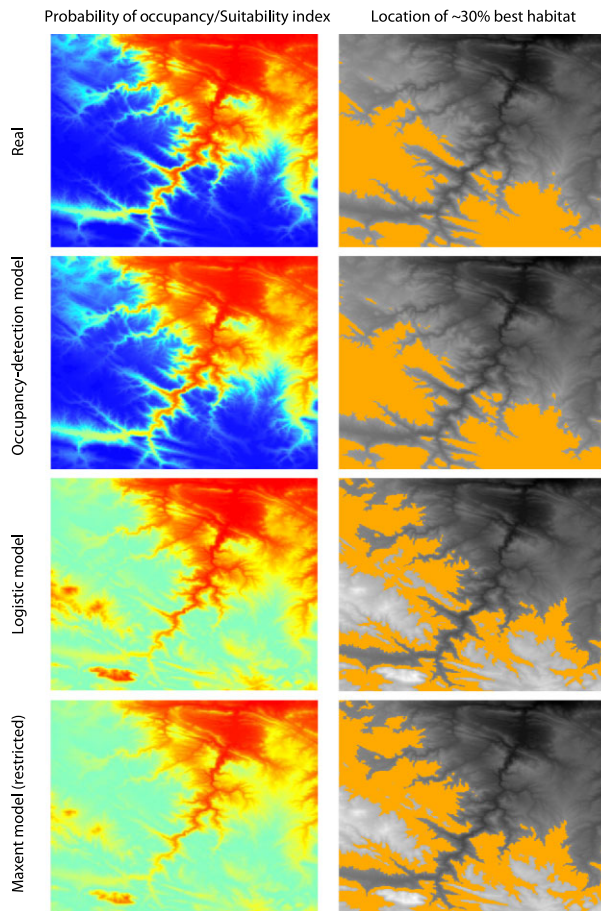
Probability of occupancy/Suitability index   Location of ~30% best habitat

Real

Occupancy–detection model

Logistic model

Maxent model (restricted)

Figure 4 Real and estimated distribution of the virtual species under a simulated scenario with detectability negatively correlated to occupancy and a steep transition in occupancy (see parameter values in the main text). The left column shows maps of occupancy and the right column the (approximated) 30% of pixels with the greatest occupancy (in orange), with elevation (greyscale) as background. Maxent estimates represent a suitability index, not a probability. AUC values were 0.93, 0.93, 0.86 and 0.88 for perfectly calibrated, occupancy–detection, logistic and Maxent models respectively.

Estimating distributions in geographical space

The results for the scenarios of negative correlation (Fig. 4) illustrate the potential impact that disregarding the effect of imperfect detection can have in the estimation of the distribution of a species (see Appendix S3 for other scenarios). As shown in Fig. 1 for environmental space, when the detectability was not taken into account both logistic and Maxent models interpreted the small number of detections at high elevation as due to lower occupancy/suitability. Overall, the logistic model underestimated the probability of occupancy. The occupancy–detection model on the other hand was able to tease apart the effect of detectability at higher elevations, providing an accurate estimate of the species distribution. In this example the areas selected as the top 30% best habitat by the logistic and Maxent models only contained 30% and 37% respectively of the real top 30%.

Table 2 Summary of the impacts that disregarding imperfect detection can have on the performance of species distribution models as a function of the structure of the detection process and its correlation with the occupancy process.

	Detectability structure			
		Varying		
Performance aspect	Constant	Positive correlation	Negative correlation	Independent
Discrimination	✓	✓	✗	✗
Calibration	✗	✗	✗	✗

Symbol ✗ represents that imperfect detection affects performance, and ✓ that it does not. This applies to both presence–absence and presence–background methods. Note that presence-background methods such as Maxent produce a suitability index rather than a probability and are thus not necessarily calibrated even assuming perfect detection.

DISCUSSION

The impact of imperfect detection in SDMs and their applications

We have shown that imperfect detection affects SDM calibration and that its impact on model discrimination depends on the relationship between detectability and occupancy (Table 2). Our demonstration utilized logistic regression and Maxent as examples of presence–absence and presence–background techniques, but our general results are applicable to other approaches that do not account for imperfect detection. This includes Poisson point process models for which correspondence with Maxent has been shown (Fithian & Hastie, 2013; Renner & Warton, 2013). Despite assertions to the contrary (e.g. Hirzel *et al.*, 2002; Gibson *et al.*, 2007), presence–background methods are not immune to imperfect detection (Kéry *et al.*, 2010; Elith *et al.*, 2011; Dorazio, 2012; Yackulic *et al.*, 2013). Indeed, they are essentially equally as affected as presence–absence methods, as our results illustrate. Although avoiding the use of potential false absence records, the negative influence of imperfect detection remains a problem for presence–background methods because, under spatially varying imperfect detection, the set of presence records is not an unbiased sample of all occupied locations. Importantly, we show that imperfect detection can also be a problem when detection probability depends on a covariate that is *independent* from those driving occupancy. If considered as a candidate explanatory variable, the model can wrongly identify the detection covariate as an occupancy covariate, thus resulting in poor inference and predictions (cf. Dorazio, 2012, who assesses through mathematical proof the case where the detection covariate is not considered in the modelling).

Whether the influence of imperfect detection is an issue in practice is also dependent on the application. We have seen that if detection is not perfect but can be assumed to be constant or positively correlated with occupancy the ability to discriminate

between presence and absence locations is not impaired. Hence, if the objective is for instance to prioritize areas by ranking sites in terms of habitat quality for a single species (e.g. Lahoz-Monfort *et al.*, 2010), then the output of the SDM will still be useful despite the model not being well calibrated. Such ranking can be relevant when identifying areas for the reintroduction of captive-bred endangered species, for guiding the search for an unknown population of a species or for anticipating the establishment of invasive species. A misleading indication of the relative habitat quality of different locations could have disastrous consequences for these conservation activities. We have shown that this can happen if detectability is negatively correlated with occupancy, or if it depends on other environmental variables.

When the application of the SDM requires estimating the actual probabilities of occupancy, imperfect detection will be an issue regardless of its structure as it leads to underestimation. For instance, when conducting spatial prioritization of investment for multiple species (as in systematic conservation planning), poor calibration can lead to incoherent trade-offs between the locations of habitat for different species. Knowledge about actual probabilities might also be required, for instance as an input for survey design and associated conservation strategies, such as the optimal surveillance of invasive species (Hauser & McCarthy, 2009) or for mapping actual occupancy for the purposes of impact assessment or analysis of population viability (Southwell *et al.*, 2008). Note that, for these types of applications, presence–background techniques such as Maxent, which produce a relative index rather than a probability, will not be suitable regardless of detectability, unless there are other means of calibrating the models.

Importance of assessing model performance with respect to model objectives

Performance has to be assessed with respect to objectives. We have emphasized that, since our objective is to model the distribution of species, an assessment of the impacts of imperfect detection on SDMs needs to be based on how well different techniques are able to characterize and predict a species' distribution. Using simulations that capture a range of plausible scenarios likely to be encountered with empirical datasets, we have shown that accounting for imperfect detection can provide significant improvement in terms of model prediction and discrimination capacity. Previous studies (Rota *et al.*, 2011; Comte & Grenouillet, 2013) that suggested little value in modelling imperfect detection were based on assessing performance in modelling detections but, as we have argued, this does not reflect the aim of species distribution modelling. We acknowledge that modelling detections could potentially be of interest in some specific situations (e.g. identifying good spots for hunting or wildlife viewing), but then we would no longer be dealing with species *distribution* models but rather with species *detection* models. Even if this were the aim, an occupancy–detection model would produce estimates of where the species is detected, with the added benefit of providing the picture of where the

species is most likely to occur and, if there is spatial variation in detectability, where it is most likely to be detected if present.

Assessing true predictive performance of a SDM using empirical observational data that may be affected by imperfect detection is not possible because we only have imperfect knowledge about the partially observable occupancy process (e.g. Zipkin *et al.*, 2012). Model evaluation based on detections can still help identify models with poor fit because models that are unable to capture the observable process cannot be expected to capture well the latent process of occupancy. For instance, MacKenzie & Bailey (2004) propose a goodness-of-fit test to detection data for occupancy–detection models. However, goodness-of-fit to detection data should not be used for selection between models that account for detectability and models that disregard it, as they may both fit the observable process well, while only those accounting for detectability will provide good estimates of occupancy.

Accounting for detectability and data needs

Simply increasing the size of the training sample (i.e. the number of sampled sites) does *not* solve the problem of imperfect detection. For instance, a ten-fold increase in the number of training sites in our simulations (to 4000) resulted in practically the same median AUC values, e.g. for the scenario 'steep' with 'negative correlation', we obtained 0.94 (0.93–0.95) with the occupancy–detection model, 0.87 (0.84–0.89) with logistic regression and 0.87 (0.85–0.89) with Maxent. In fact, the prediction of environmental relationships can be even more misleading since the precision increases while the bias is maintained, making us more 'certain' about an incorrect inference.

There are essentially two options to avoid the problems of imperfect detection: either to expend enough effort at surveyed sites so that detection can be assumed perfect or to collect survey data in a way that allows the detection process to be modelled. It is important to note that the latter does not necessarily imply greater survey effort. In many instances changing the observation and recording practices only slightly ensures that collected data are informative about detectability. Examples include recording at each site the time to first detection (Garrard *et al.*, 2008), inter-detection intervals (Guillera-Arroita *et al.*, 2011), the individual detections by multiple independent observers (MacKenzie *et al.*, 2006) or the observations gathered in multiple visits (MacKenzie *et al.*, 2002). It is also worth noting that there are currently compiled data sets that can be amenable to occupancy–detection modelling even if they were not initially designed for this purpose (Kéry *et al.*, 2010).

Although modelling the detection process does not necessarily require more data, for a given level of sampling effort per site more sampling sites are needed to obtain good occupancy predictions when detectability is low. This need for increased sample size is not surprising: imperfect detection blurs the information on species occupancy, making it harder to estimate these two 'entangled' processes separately. Since, as stated above, the alternative approach to reduce the detection problem is to increase the survey effort per site, there is a trade-off in the

allocation of survey effort between number of sites and effort per site (MacKenzie & Royle, 2005; Guillera-Arroita *et al.*, 2010; Guillera-Arroita & Lahoz-Monfort, 2012). It is crucial to make sure that enough data are collected so that the predictions of SDMs are informative for their application.

When modelling observational data using modelling methods based on sampling units (e.g. quadrats or grid cells in a surveyed landscape), the parameters depend on the spatial resolution of the data. On that basis, some researchers have criticized the use of quadrat or grid-based modelling methods to predict species distributions, preferring instead point-process models (Warton & Shepherd, 2010) which are invariant to spatial resolution and can be fitted to presence-only data without aggregating the observations into spatial pixels of arbitrary size. However, a key advantage of working with sampling units is that it provides a context in which to model the detection process in occupancy–detection models. Instead, point process methods do not directly allow us to account for imperfect detection (Warton & Shepherd, 2010); auxiliary detection data are required given that presence-only data do not contain information about the detection mechanism.

CONCLUSION

Species distribution modelling is a very valuable tool for ecology, management and conservation. With the development of statistical methods, the availability of software packages, advances in computational power and the increased accessibility of survey data, SDMs are becoming routinely used to inform conservation action. To date there has been little acknowledgement of the potentially pervasive effects of imperfect detection in the predictive ability of SDMs, which can have serious implications for the effectiveness of applications that rely on their predictions. The need to invest in good data is a keystone for meaningful species distribution modelling (Lobo, 2008). Ideally, one would dedicate sufficient survey effort to ensure that detection is perfect. Since this will seldom be practical or efficient, recording data in ways that allow the modelling of the detection process should be standard practice in future surveys. When such data are not available, diligent consideration and reporting of the possible impacts of imperfect detection, including how it is likely to bias inference and prediction, should be considered a minimum standard of good species distribution modelling practice. Unless there are solid grounds to expect imperfect detection not to be a problem, modellers should recognize that their SDM could just be capturing where the species is more likely to be detected and should thoughtfully establish whether such information is still meaningful for their purposes.

ACKNOWLEDGEMENTS

REFERENCES

Altwegg, R., Wheeler, M. & Erni, B. (2008) Climate and the range dynamics of species with imperfect detection. *Biology Letters*, **4**, 581–584.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

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

Comte, L. & Grenouillet, G. (2013) Species distribution modelling and imperfect detection: comparing occupancy versus consensus methods. *Diversity and Distributions*, **19**, 996–1007.

Conkin, J.A. & Alisauskas, R.T. (2013) Modeling probability of waterfowl encounters from satellite imagery of habitat in the central Canadian Arctic. *Journal of Wildlife Management*, **77**, 931–946.

Dorazio, R.M. (2012) Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*, **68**, 1303–1312.

Elith, J., Graham, C.H., Anderson, R.P. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

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

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

Fithian, W. & Hastie, T. (2013) Finite-sample equivalence in statistical models for presence-only data. *Annals of Applied Statistics*. In press.

Garrard, G.E., Bekessy, S.A., McCarthy, M.A. & Wintle, B.A. (2008) When have we looked hard enough? A novel method for setting minimum survey effort protocols for flora surveys. *Austral Ecology*, **33**, 986–998.

Gibson, L., Barrett, B. & Burbidge, A. (2007) Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions*, **13**, 704–713.

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

Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution*, **3**, 860–869.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J.T. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131–139.

Guillera-Arroita, G., Morgan, B.J.T., Ridout, M.S. & Linkie, M. (2011) Species occupancy modeling for detection data collected along a transect. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**, 301–317.

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.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Hauser, C.E. & McCarthy, M.A. (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters*, **12**, 683–692.

Hijmans, R.J., Phillips, S.J., Leathwick, J.R. & Elith, J. (2013) *Dismo: species distribution modeling*. R package version 0.8-11. Available at: http://CRAN.R-project.org/package=dismo (accessed 18 November 2013).

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.

Jiménez-Valverde, A. (2012) Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, **21**, 498–507.

Jiménez-Valverde, A., Acevedo, P., Barbosa, A.M., Lobo, J.M. & Real, R. (2013) Discrimination capacity in species distribution models depends on the representativeness of the environmental domain. *Global Ecology and Biogeography*, **22**, 508–516.

Kearney, M.R., Wintle, B.A. & Porter, W.P. (2010) Correlative and mechanistic models of species distribution provide congruent forecasts under climate change. *Conservation Letters*, **3**, 203–213.

Kéry, M. (2002) Inferring the absence of a species: a case study of snakes. *Journal of Wildlife Management*, **66**, 330–338.

Kéry, M. (2011) Towards the modelling of true species distributions. *Journal of Biogeography*, **38**, 617–618.

Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216.

Kéry, M., Gardner, B. & Monnerat, C. (2010) Predicting species distributions from checklist data using site-occupancy models. *Journal of Biogeography*, **37**, 1851–1862.

Kéry, M., Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2013) Analysing and mapping species range dynamics using occupancy models. *Journal of Biogeography*, **40**, 1463–1474.

Kujala, H., Moilanen, A., Araújo, M.B. & Cabeza, M. (2013) Conservation planning with uncertain climate change projections. *PLoS One*, **8**, e53315.

Lahoz-Monfort, J.J., Guillera-Arroita, G., Milner-Gulland, E.J., Young, R.P. & Nicholson, E. (2010) Satellite imagery as a single source of predictor variables for habitat suitability modelling: how Landsat can inform the conservation of a critically endangered lemur. *Journal of Applied Ecology*, **47**, 1094–1102.

Lobo, J.M. (2008) More complex distribution models or more representative data? *Biodiversity Informatics*, **5**, 14–19.

Lobo, J.M., Jiménez-Valverde, A. & Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**, 145–151.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*, 2nd edn. Chapman and Hall/CRC, London.

MacKenzie, D.I. & Bailey, L.L. (2004) Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, **9**, 300–318.

MacKenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., 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., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Academic Press, New York.

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.

Monk, J. (2013) How long should we ignore imperfect detection of species in the marine environment when modelling their distribution? *Fish and Fisheries*. doi: 10.1111/faf.12039.

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

Peterson, A.T., Soberon, J., Pearson, R.G., Anderson, R.P., Martinez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011) *Ecological niches and geographic distributions*. Princeton University Press, Princeton, NJ.

Phillips, S.J. & Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

R Development Core Team (2011) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Renner, I.W. & Warton, D.I. (2013) Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics*, **69**, 274–281.

Rota, C.T., Fletcher, R.J., Evans, J.M. & Hutto, R.L. (2011) Does accounting for imperfect detection improve species distribution models? *Ecography*, **34**, 659–670.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical modeling and inference in ecology*. Academic Press, Amsterdam.

Southwell, D.M., Lechner, A., Coates, T. & Wintle, B.A. (2008) The sensitivity of population viability analysis to uncertainty about habitat requirements: implications for the management of the endangered southern brown bandicoot. *Conservation Biology*, **22**, 1045–1054.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing

bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the 'pseudo-absence problem' for presence-only data in ecology. *Annals of Applied Statistics*, **4**, 1383–1402.

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Grant, E.H.C. & Veran, S. (2013) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.

Yoccoz, N.G., Nichols, J.D. & Boulinier, T. (2001) Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution*, **16**, 446–453.

Zipkin, E.F., Campbell Grant, E.H. & Fagan, W.F. (2012) Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. *Ecological Applications*, **22**, 1962–1972.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix S1 Plots of simulated scenarios (ecological relationships).
Appendix S2 AUC (Area under the ROC curve) values for the perfectly calibrated models.
Appendix S3 Projecting into geographical space: simulating from a real landscape.
Appendix S4 Other simulation results.

BIOSKETCHES

Gurutzeta Guillera-Arroita is currently a member of the Environmental Decisions Group and Research Fellow at the University of Melbourne. She is interested in the development and application of statistical models to the study of wildlife populations, with a focus on optimal study design, as well as on optimal decision-making for conservation problems.

José J. Lahoz-Monfort is currently a member of the Environmental Decisions Group and Research Fellow at the University of Melbourne. His interests include the modelling of animal demography, population dynamics and species distributions in the context of applied ecology and biodiversity conservation.

Brendan Wintle is Associate Professor of Conservation Ecology at the University of Melbourne and an ARC Future Fellow. He is interested in ecological modelling and in characterizing and dealing with uncertainty in environmental decisions.

Editor: Peter van Bodegom