A comparison of Maxlike and Maxent for modelling species distributions

Cory Merow[1,2]* and John A. Silander Jr[2]

[1]*Smithsonian Environmental Research Center, 647 Contees Wharf Road, Edgewater, MD 21307-0028, USA; and* [2]*Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., Storrs, CT 06269, USA*

Summary

1. Understanding species spatial occurrence patterns and their environmental dependence is one of the fundamental goals in ecology and evolution. Often, occurrence models are built with presence-only data because absence data are unavailable. We compare the strengths and limitations of the recently developed presence-only modelling method, Maxlike, with the more widely used Maxent.

2. In spite of disparities highlighted by the developers of Maxlike and Maxent, we show approximate formal relationships between the parameters of Maxlike and Maxent for two scenarios to illustrate their similarity. Using case studies based on real and simulated data, we show how these similarities manifest in practice.

3. We find more similarities than differences between Maxlike and Maxent, including coefficient values, predicted spatial distributions, similarity to presence–absence models, predictive performance and ranking the predicted suitability of cells. Maxlike reliably predicted absolute occurrence probabilities for very large data sets on landscapes where occurrence probability approximately spanned [0,1]. For smaller data sets, the uncertainty in predicted occurrence probability by Maxlike was very large, due to the inherent limitations of presence-only data. In contrast, Maxent is constrained to predicting relative occurrence probabilities or relative occurrence rates unless it is provided with additional information from presence–absence data. Both models can reliably predict relative differences in occurrence probability.

4. The choice of which model to use depends partly on sampling assumptions, which we discuss in detail. Due to limitations of presence-only data, ecologists should typically focus on interpretations relying on relative differences in occurrence probability or relative occurrence rates. We discuss how to remedy a number of concerns about the use of Maxent and how to avoid some potential pitfalls with Maxlike – particularly related to high variance predictions. We conclude that both methods are similarly valuable for understanding and predicting species' distributions in terms of relative differences in occurrence probability when the models are specified carefully.

Key-words: occurrence model, presence-only data, probability of presence, resource selection function, species distribution modelling

Introduction

Understanding species spatial occurrence patterns and their environmental dependence is one of the fundamental goals in ecology and evolution. The most common type of occurrence models correlates records of presence, and sometimes absence, with environmental covariates. While correlative occurrence models have some important shortcomings (Anderson 2012; Araújo & Peterson 2012), the limitation of available data makes them one of the few tools available for understanding spatial distributions for the majority of species on earth. Data limitations are particularly apparent for presence-only (PO) data (Graham *et al.* 2004) – composed only of presence (and not absence) records – which has lead to a variety of modelling methods to maximize inference (Manly, McDonald & Thomas 1993; Lele & Keim 2006; Phillips, Anderson & Schapire 2006; Royle *et al.* 2012). The most popular PO model has been

Maxent (>2000 citations; Phillips, Anderson & Schapire 2006), which was originally developed in a machine learning framework but has been recently derived via maximum likelihood (Halvorsen 2012) and related to a variety of generalized linear models (Aarts, Fieberg & Matthiopoulos 2012; Fithian & Hastie 2013; Renner & Warton 2013). Recently, Royle *et al.* (2012) advocated an occurrence model for PO data, informally called *Maxlike*, to remedy issues that they found with Maxent. Royle *et al.* show that Maxlike is capable of estimating absolute *occurrence probability* (OP; the probability that the species is present in a grid cell), whereas Maxent cannot. In particular, Royle *et al.* focus criticism on the use of Maxent's logistic output to predict OP, with which we concur. Here, we illustrate the formal and practical relationship between Maxlike and Maxent under different sampling assumptions and a more appropriate application of Maxent.

The typical data for PO occurrence models consist of three components: (i) a collection of point locations where a species has been recorded, (ii) a gridded landscape that the modeller

*Correspondence author. E-mail: cory.merow@gmail.com

considers available to the species of interest (i.e. the species could migrate to all locations in that landscape), and (iii) environmental covariates for each of the grid cells. The point locations are typically aggregated to the grid cell in which they occur. These models assume that grid cells have been randomly sampled for presence, and thus, records occur in proportion to the species' habitat usage preferences. Regardless of whether 1 individual or 100 individuals are observed, a 1 is recorded. Absences (0's) are unknown. *Background samples* are selected randomly from the gridded landscape to compare against the presence cells and determine whether environmental conditions are used disproportionately to their availability (Keating & Cherry 2004; Lele & Keim 2006). The goal of the analysis is to determine whether environmental conditions are used (presences) disproportionately to their availability (background).

Though it may seem appealing to understand species' occurrence patterns in terms of OP, its definition poses some practical limitations. The OP of a grid cell describes the proportion of (hypothetical) cells with equivalent environmental conditions where a species is expected to be observed. This definition of OP is not purely biological, as we might intend; the OP depends on the size of the cell (larger cells should have larger OP) and is typically confounded with sampling effort (more search effort would lead to a larger probability of observation). Alternative models have been built to accommodate both of these limitations (e.g. Kery, Gardner & Monnerat 2010; Chakraborty *et al.* 2011; Keil *et al.* 2013); however, they require information on abundance or measures of sampling effort, which are typically unknown. Here, we focus on the more common occurrence models that ignore these issues, although these limitations are pertinent for model interpretation.

In this article, we begin by discussing the way in which Maxlike and Maxent model species occurrence. We show that Maxent is better suited to model relative abundance under different sampling assumptions than those typically made for modelling OP. We then illustrate the similarity between the methods by showing two scenarios where approximate analytic relationship between the slope parameters of each model hold. Consequently, we find more similarities than differences between Maxlike and Maxent. In the discussion, we highlight pitfalls of practical applications of both models and how they can be addressed.

Model definition

MODELLING OCCURRENCE PROBABILITY WITH MAXLIKE AND MAXENT

To formalize the differences between Maxlike's and Maxent's predictions, it is helpful to refer to their common likelihood function. Using $y_i = 1$ to denote a presence at grid cell x_i, and $\psi(y_i = 1|x_i, \beta_0, \beta)$ to denote OP, the likelihood for Maxlike is given by Royle *et al.* (2012):

$$L(\beta) = \Pi_{i=1}^{N} \frac{\psi(y_{i=1}|x_i, \beta_0, \beta)}{\sum_{x \in B} \psi(y_{i=1}|x_i, \beta_0, \beta)} \qquad \text{eqn 1}$$

where N is the total number of presences, B is the set of background cells, β_0 is an intercept parameter, and β is the vector of slope coefficients associated with environmental covariates. The numerator describes the likelihood at presence cells while the denominator describes the likelihood at background cells. Often, background cells are taken as a random sample of cells on the landscape where the model is fit (Lele & Keim 2006; Lele 2009; Royle *et al.* 2012).

The entropy maximization performed by Maxent turns out to be equivalent to maximizing the likelihood in equation 1 under certain assumptions. The maximum entropy distribution (described in detail below) can be obtained by maximizing the likelihood of a Gibbs distribution, given by $q(\beta) = e^{\beta z(x)} / \sum_B e^{\beta z(x)}$, where $z(x)$ is a vector of environmental covariates in cell x (Della Pietra *et al.* 1997; Phillips *et al.* 2004). Given this equivalence, and comparing equation (1) with the Gibbs distribution, it is apparent that Maxent's model can be obtained by maximizing the likelihood in equation (1) by making the parametric assumption that $\psi(y = 1|x, \beta) = e^{\beta z(x)}$. Note that β_0 is missing from this equivalence, a point that we return to below. Operationally, Maxent optimizes a penalized version of the likelihood in eqn 1 (Phillips & Dudik 2008; Merow, Smith & Silander 2013).

The difference between Maxlike and Maxent derives from the choice of link functions used for ψ. Maxlike uses a logit-linear model:

$$\psi(y_i = 1|x_i, \beta_0, \beta) = e^{\beta_0 + \beta z(x_i)} / (1 + e^{\beta_0 + \beta z(x_i)}) \qquad \text{eqn 2}$$

Maxent uses a loglinear model:

$$\psi(y_i = 1|x_i, \beta) = e^{\beta z(x_i)} \qquad \text{eqn 3}$$

The important difference between these functions is the intercept, β_0, which differentiates their ability to predict OP. Using the loglinear model, Maxent cannot estimate an intercept because it cancels from the likelihood in equation 1 (Royle *et al.* 2012). The intercept effectively defines the expected prevalence across a landscape, which attaches a scale to the relative differences defined by the slope parameters (Fig. 1; Hastie & Fithian 2013). Prevalence is defined as $\psi(y = 1) = \sum_B \psi(y_i = 1|x_i)$, that is, the proportion of occupied cells, or the average OP across the landscape; hence, large values of the intercept increase the prevalence.

Maxlike

Maxlike uses a logit-linear model for two reasons. First, a logistic model ensures that predicted OP lies on [0,1], whereas Maxent's loglinear model does not. Secondly, because no two logistic functions are exactly proportional, the intercept in equation (2) is estimable from the likelihood in equation (1). Hastie & Fithian (2013) point out that the lack of proportionality that enables Maxlike to differentiate the likelihoods for different values of the intercept may be very slight. Indeed, our demonstrations below show that this leads to high variance estimates of the intercept.
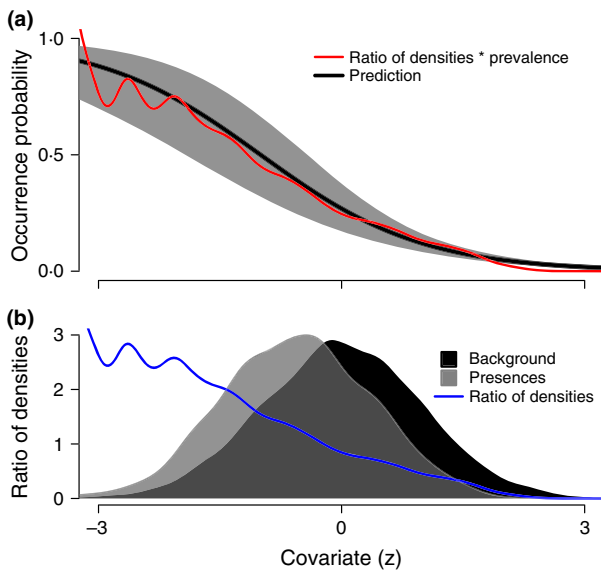
Fig. 1. Graphical illustration of Maxlike's estimation of prevalence from presence-only data. Values of an environmental covariate (z) were simulated at 10 000 background cells, and 1000 presences were sampled with probability $\psi(z) = -1-1*z$. (a) The black line (grey interval) shows Maxlike's median (95% interval) predicted OP over 500 replicate presence samples. (b) The black (grey) probability density shows the empirical distribution of z at background (presence) cells. The red and blue lines illustrate how Maxlike estimates prevalence (i.e. the intercept in eqn 2). The blue line shows the empirical ratio between the two densities in (b) while the red line shows this same ratio multiplied by the estimated value of prevalence (eqn 4). Maxlike chooses the value of prevalence by which to multiply the empirical ratio (blue) to make it as similar as possible (red) to a logistic curve (black). Note the wide range of predicted OP (grey 95% interval) for negative values of the covariate that can result from even a simple univariate model with a large sample size.

It may seem counterintuitive that prevalence is estimable from PO data. To determine the proportion of occupied cells, it would seem the proportion of unoccupied cells must also be known. To get around this data limitation, Maxlike takes advantage of the definition of OP: the proportion of occupied cells is equivalent to the average true OP across the landscape. Of course, the true OP is unknown, so Maxlike assumes that the average *predicted* OP across the landscape approximates the average true OP (denominator in eqn 1).

To illustrate the connection between Maxlike and data shown in Fig. 1, it is helpful to rewrite $\psi(y = 1|x)$, which describes OP in geographic space, as $\psi(y = 1|z(x))$, which describes OP in terms of environmental covariate space. Here, $z(x)$ denotes the set of environmental covariates in cell x (Aarts, Fieberg & Matthiopoulos 2012; Royle *et al.* 2012). Using Bayes' rule, the OP can be rewritten as (Phillips *et al.* 2009),

$$\psi(y = 1|z(x)) = \frac{\pi(z(x)|y = 1)}{\pi(z(x))}\psi(y = 1) \qquad \text{eqn 4}$$

where $\pi(z(x)|y = 1)$ is the empirical probability density of the environmental predictors at presence cells and $\pi(z(x))$ is the empirical probability density of the environmental predictors at background cells. These empirical probability densities can be thought of simply as frequency histograms of a predictor, for example mean annual precipitation (Fig. 1b). Figure 1 graphically illustrates how Maxlike predicts OP for a model with a single environmental covariate; OP is a model for the ratio of the density at presence cells (grey) to background cells (black), rescaled by the prevalence. The blue line gives the exact, empirical ratio between the densities, while the red line shows the same ratio multiplied by the estimated value of the prevalence. Maxlike is able to estimate the intercept by estimating a value for the prevalence that causes the red line to be as close as possible to a logistic curve (black line). Hence, Maxlike's parametric assumption is critical for determining the intercept and hence the predicted prevalence.

Maxent

Maxent's loglinear model and its inability to estimate an intercept can be understood by its connection to inhomogeneous poisson process (IPP) models (cf. Warton & Shepherd 2010; Chakraborty *et al.* 2011) although other derivations are possible (Phillips, Anderson & Schapire 2006; Halvorsen 2012; Merow, Smith & Silander 2013). IPPs provide a general framework linking a variety of occurrence modelling strategies, including Maxent, discretized Poisson count models, naïve logistic regression (using the background as absences) and weighted logistic regression (Aarts, Fieberg & Matthiopoulos 2012; Fithian & Hastie 2013; Renner & Warton 2013). Although our discussion below focuses on comparison of Maxlike with Maxent, it extends more generally to a variety of generalized linear models under appropriate assumptions about discretization of space, distributions and integration techniques.

Rather than assuming random sampling of space, as Maxlike does, IPPs assume that presence records represent a random sample of individuals, and thus, records occur in proportion to population density ('density estimation'; Dudik, Phillips & Schapire 2004; Baddeley *et al.* 2010). With this assumption, the goal is to predict the density of presence records at point locations in continuous space (as opposed to grid cells) using an intensity function, denoted $\lambda(x)$ (Warton & Shepherd 2010). $\lambda(x)$ is often called an occurrence rate and has units of number of records per unit area (Fithian & Hastie 2013). Integrating $\lambda(x)$ over a region A returns the expected number of presence records in A and is given by $\Lambda(A) = \int_A \lambda(x)\,dx$. Integrating $\lambda(x)$ over the entire landscape returns the expected total number of presence records. The IPP's occurrence rate derives from the continuous limit of a Poisson count model for increasingly fine discretizations of the landscape (Fithian & Hastie 2013). If N(A) denotes the number of observations in area A, it is distributed as (Fithian & Hastie 2013):

$$N(A) \sim Poisson(\Lambda(A)) \qquad \text{eqn 5a}$$

As A approaches zero, $\Lambda(A)$ approaches $\lambda(x)$. $\lambda(x)$, is often modelled as a loglinear function (Warton & Shepherd 2010):

$$\log(\lambda(x)) = \beta_0 + \boldsymbol{\beta}z(x) \qquad \text{eqn 5b}$$

Notably, the IPP estimates an intercept.

Maxent differs in two ways from the IPP: it uses a discretized landscape rather than a continuous one and it assumes that the total number of records is fixed rather than random. To achieve this, Maxent uses a discrete Poisson model in cell x_i:

$$N(x_i) \sim Poisson(\lambda(x_i)) \qquad \text{eqn 6a}$$

with

$$\log(\lambda(x_i)) = \boldsymbol{\beta} z(x_i) \qquad \text{eqn 6b}$$

The first difference between these equations (5) and (6) is that presence records, which occur across a continuous landscape, are now aggregated to the grid cell x_i. The second difference is that Maxent does not estimate the intercept (Aarts, Fieberg & Matthiopoulos 2012). The intercept in the IPP (eqn 5a) determines the total number of presence records, which is typically not of interest (Fithian & Hastie 2013). Often, the counts are proportional, but not equal, to abundance, and hence, one is restricted to modelling the *relative occurrence rate* (ROR; Fithian & Hastie 2013; Renner & Warton 2013), as Maxent does. To understand the distinction between abundance and ROR, consider the following example (adapted from Fithian & Hastie 2013); for the Carolina wren data shown in Fig. 2, 1506 presences were observed. If twice as many presences were observed, should that indicate that there are twice as many wrens or twice as many birders looking for wrens? Because all individuals are not observed in any subset of cells, the best we can do is model relative differences in occurrence rates between cells. The ROR is typically reported as a density function given by,

$$\pi(x_i) = \lambda(x_i) / \sum_{x \in B} \lambda(x_i) = e^{\boldsymbol{\beta} z(x_i)} / \sum_{x \in B} e^{\boldsymbol{\beta} z(x_i)} \qquad \text{eqn 7}$$

The ROR, $\pi(x)$, is known as the software's *raw output*. Given that an individual was observed, $\pi(x)$ describes the relative probability that the individual derived from each cell on the landscape (Phillips, Anderson & Schapire 2006; Merow, Smith & Silander 2013). Alternatively, if $\pi(x)$ is interpreted in environmental space as in equation 4 ($\pi(z(x))$), it describes the predicted probability density of presences in environment z (Fig. 1).

From this derivation of Maxent's prediction, it seems clear that Maxent is better suited for predicting ROR than OP. By comparison of Maxent's raw output (eqn 7) to the likelihood function in equation 1, it is apparent that $\lambda(x)$ is equivalent to $\psi(y = 1|x)$ (Royle *et al.* 2012). However, interpreting Maxent's prediction of $\lambda(x)$ as OP is problematic for two reasons: $\lambda(x)$ can be greater than unity and $\lambda(x)$ derives from the assumption of random sampling of individuals, not the random sampling of space typically associated with OP (Royle *et al.* 2012). The unboundedness on [0,1] illustrates why additional assumptions are needed to convert these predictions to OP in the Maxent software package (Phillips & Dudik 2008). Due to problems with these assumptions (which lead to so-called *logistic output*, discussed in Section V.A) we focus only on Maxent's raw output, for the remainder of the paper.

Nonetheless, further accommodations have been made to interpret Maxent's output in terms of OP. To handle the issue of sampling design (individuals vs. space), Maxent duplicate observations in a grid cell to approximate a random sample of space. Conditional on the acceptability of this aggregation and thinning of the data (see discussion in Aarts, Fieberg & Matthiopoulos 2012), raw output can be interpreted as OP (ψ in eqn 3) rescaled by a constant (the sum of predictions over all background cells). Hence, raw output is proportional to OP, and we refer to it as *relative OP* when treating the data as a random sample of space. Relative OP describes how cells compare to one another, with the ratio of relative OP equal to the ratio of OP between two cells (Lele & Keim 2006).
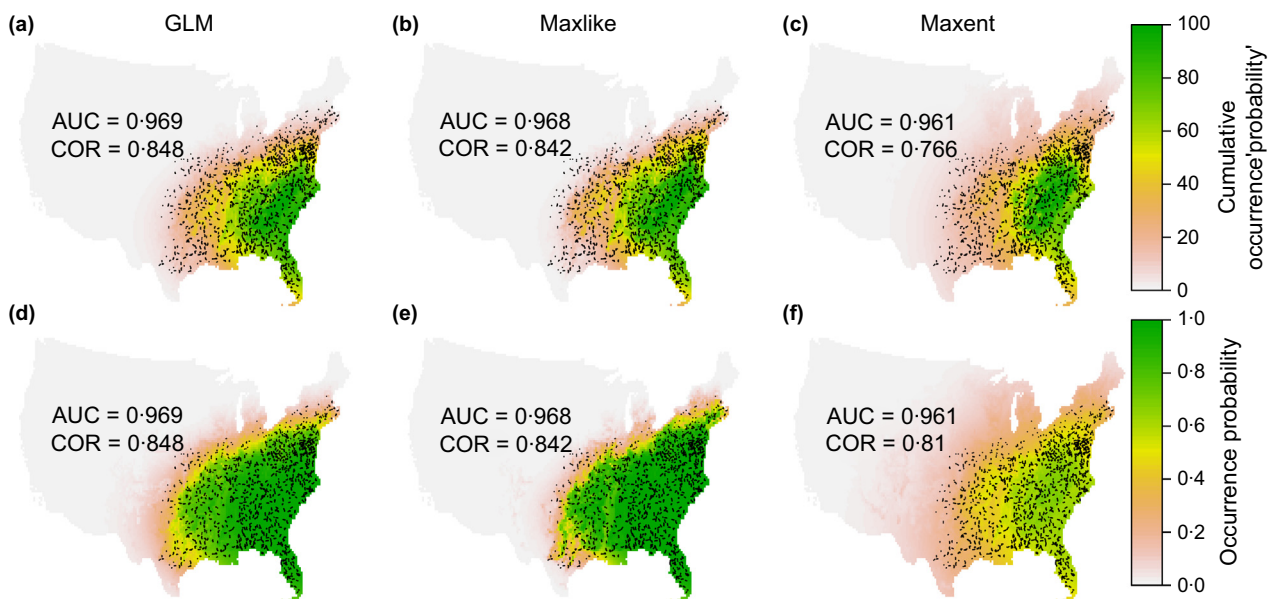

Fig. 2. Comparison among predictions from a GLM using PA data and Maxlike and Maxent using PO data. Panels (a–c) show the models described in this paper for relative OP on the cumulative output scale, while (d–f) are reproduced from Royle *et al.*' (2012) models for absolute OP. AUC and COR are calculated with respect to PA data. All models of relative OP predicted very similar spatial patterns.

LINKING MAXLIKE AND MAXENT

To summarize the previous two sections, Maxent's output can be interpreted in one of two ways, depending on sampling assumptions. If data are assumed to be a random sample of space, we can predict relative OP with Maxent, but not OP. If data are assumed to be a random sample of individuals, Maxent can predict ROR. Each of these interpretations is formally linked to Maxlike under the circumstances discussed below.

First, we consider the case where Maxent is used to model relative OP. In the limit of decreasing cell size on the discretized landscape, the parameters of logit-linear and loglinear models are asymptotically equivalent (Baddeley *et al.* 2010). As cell size decreases, the value of OP (denoted by ψ) in a given cell decreases. Maxlike models logit(ψ) = log($\psi/(1-\psi)$)$\approx$log(ψ), where approximate equivalence is achieved for small ψ. Because Maxent models log(ψ) explicitly, the models are equivalent in the limit of a high-resolution gridded landscape. This is not surprising, because an exponential function looks reasonably similar to a logistic function until the logistic function asymptotes.

Secondly, we consider the case where Maxent is used to model ROR. The probability that a cell is occupied is equal to the probability that the number of counts there is greater than zero. The probability of non-zero counts is equal to 1 minus the probability of zero counts:

$$\psi(y_i = 1|x_i) = \Pr(N(x_i) > 0) = 1 - \Pr(N(x_i) = 0) \qquad \text{eqn 8}$$

For the Poisson model in equation 5, the probability of k observations in a cell is given by $\Pr(N((x_i) = k) = \lambda^k exp(-\lambda(x_i))/k!$. Hence, the probability of zero observations is $Pr(N = 0) = exp(-\lambda(x_i))$. Using this result with a general loglinear model for $\lambda(x_i)$ (*e.g.* eqn 5b), we obtain,

$$\psi(y_i = 1|x_i) = 1 - \exp(\lambda(x_i)) = 1 - \exp(-\exp(\beta_0 + \boldsymbol{\beta}\mathbf{z}(x_i))) = \text{cloglog}^{-1}(\beta_0 + \boldsymbol{\beta}\mathbf{z}(x_i)) \qquad \text{eqn 9}$$

This result shows that the slope coefficients ($\boldsymbol{\beta}$) estimated by a model for ROR, using count data, are closely related to those obtained from a model for OP, using presence–absence data, if a complementary log–log link function were used (Baddeley *et al.* 2010). To the extent that the logit link (used by Maxlike) is a good approximation to the complementary log–log link (which is also preferred on theoretical grounds; Baddeley *et al.* 2010), and when the same method of parameter estimation is used, the slope coefficients produced by Maxlike and a model of ROR from Maxent should be approximately equivalent (as seen in Table 1).

Equation (9) highlights a general connection between models of occurrence and abundance that extends from Maxent to IPPs. If an IPP is used to obtain $\lambda(x_i)$, it is possible to estimate an intercept (conditional on the presences being a random sample of individuals). Equation (9) shows that one can move freely between estimates of ROR and OP. However, the IPP intercept estimate does not account for the possibility of missing presence points; hence, estimates of the intercept (and prevalence) may be biased low. These missing presences prevent us

Table 1. Comparison of coefficients among models for the wren data shown in Fig. 2a–c. All models exhibit the same qualitative pattern of sign and relative magnitude of coefficients. We do not expect their parameters to be identical given the models' different functional forms

Coefficient	GLM	Maxlike	Maxent
Intercept	−1·065	−1·698	NA
Latitude	−5·049	−6·671	−4·124
Latitude^2	−3·762	−5·046	−2·370
Longitude	6·409	8·173	7·621
Longitude^2	−2·051	−1·159	−2·418
% Deciduous	0·704	1·288	1·286
% Deciduous^2	−0·136	−0·236	−0·635
% Mixed Forest	−0·184	−0·224	−0·080
% Mixed Forest^2	0·025	0·021	0·000

from modelling occurrence rate, and limit us to ROR, as discussed above. In contrast, Maxlike does not appear to suffer from the influence of missing presences but instead leverages off the departure from loglinearity to estimate the intercept (which is criticized by Hastie & Fithian 2013).

These formal connections between Maxlike and Maxent show why the predictions of each model are similar, which we further illustrate with examples below.

Methods

To illustrate the similarities, strengths and weakness of Maxent and Maxlike, we use the same two-pronged approach following Royle *et al.* (2012): analysing (i) data on the distribution of the Carolina wren (*Thryothorus ludovicianus*) across the United States and (ii) the simulated data described by Royle *et al.*

Similarities between Maxent and Maxlike are apparent if their predictions are on the same relative scale and span the same range. Royle *et al.* (2012) aptly identify the problems with Maxent's logistic output (see *Discussion*), which means that we cannot compare Maxent and Maxlike predictions of OP. Instead, we degrade Maxlike's predictions of OP to relative OP by dividing by the sum of predicted OP over all cells. For display, we convert predictions to *cumulative output* (a common format for Maxent), where the value assigned to a cell is the sum of all raw values less than or equal to the raw value for that cell, rescaled to lie between 0 and 100 (Phillips, Anderson & Schapire 2006).

WREN DATA

We built models for the Carolina wren data using the same suite of features used by Royle *et al.* Because presence–absence (PA) data were available, we used logistic regression (denoted GLM) to build the 'true' model; that is, the best predictions of OP that we could achieve given absence data. We compared the GLM to Maxlike and Maxent to determine the similarity of relative OP predictions. Only linear and quadratic features were used in all models. We expect greater similarity between the GLM and Maxlike *a priori* simply because they use the same logistic functional form. Because it did not seem to relevant to make comparisons among modelling strategies while using spurious variables, we refined the variables incorporated in the model using forward and backward stepwise AIC selection on the features used by Royle *et al.* in the GLM (i.e. the 'true' model). This removed features related to the percentage coniferous forest and grassland. Using these reduced models, we also compared output from Maxlike and Maxent

models built with 50, 100, 300 and 500 presences to determine the effect of sample size on predictions. We evaluate the fit of all models to 4615 presence–absence observations use the area under the receiver–operator curve (AUC; Fielding & Bell 1997) and the point biserial correlation (COR; i.e. Pearson correlation with one binary variable; Zheng & Agresti 2000).

SIMULATED DATA

We simulated data to compare Maxlike and Maxent predictions on the relative OP scale and to explore the variability in predictions from Maxlike for different sample sizes of presences (50, 100, 200, 500, 1000 and 2000) using the same model as Royle *et al.* (2012). This model simulates the 'true' OP as a function of a covariate z (simulated from Normal(0,1)) as $logit(\psi) = -1-1*z$. We drew a single sample of z at 10 000 cells and simulated binomial sampling for presence–absence at these cells based on ψ. The absences were discarded and 500 replicated samples were drawn from the set of presences. Importantly, these replicates were drawn from the same set of simulated z values and sampled without bias, so the only source of variability in the resulting models comes from sampling the 'true' distribution of the simulated species. Models were fit using only linear features.

To understand the effect of sample size on predictions for the simulated data, we compared the distribution of mean intercept estimates for the 500 replicates for each sample size. We made a similar comparison between the ranges of predicted OP across these replicates for each cell. This range was calculated as the *maximum OP–minimum OP* for each cell over the 500 replicates. We compared the variability of Maxlike predicted OP and Maxent's predicted relative OP as a function of sample size by plotting the distribution of average percentage error within cells across the 500 replicates. Finally, we compared Maxlike's predicted OP for different values for the slope parameter (-1, -0.6 and -0.2) to explore scenarios where true OP did not span [0,1], to determine when Maxlike might struggle to estimate the intercept.

All models were built in R, v3.0.2 (R Core Team 2013) except for those fit with Maxent, v3.3.3k (Phillips, Anderson & Schapire 2006), which were subsequently accessed and analysed in R. Maxlike models

were fit with *maxlike()* in the *maxlike* package, v.0.1-3, and logistic regression models were fit with *glm()* in the *stats* package. Duplicate observations within a grid cell were removed for all models. Maxent's regularization was set to zero in order to fit an unpenalized likelihood function analogous to Maxlike's. Code for all analyses is provided in the Supplementary Materials.

Results

WREN DATA

Maxent's and Maxlike's predictions of relative OP are highly similar to one another and to the GLM based on spatial pattern and AUC and COR (Fig. 2a–c, A2 in Appendix S1). This similarity is due to our decision to model relative OP. Royle *et al.*'s differences derive from using Maxent's logistic format where differences were more apparent (Fig. 2d–f), which we suggest avoiding (Section V.C). The GLM, Maxlike and Maxent found similar environmental relationships, with similar signs and relative magnitude for all coefficients but one, which was nearly zero (Table 1).

Changing the number of presences used for training the wren model drastically altered Maxlike's spatial predictions but not Maxent's (Fig. 3). Maxlike's mean predictions are somewhat variable across sample sizes due to the inherent difficulty in estimating prevalence. In contrast, rankings of cells in these same models are consistently good (AUC > 0·847), indicating that Maxlike correctly identifies environmental relationships, that is, slope parameters. Predictions with similar (or better) AUC values are found for Maxent (Fig. 3). By focusing on relative OP (i.e. only slope parameters), Maxent is able to make consistent predictions across this range of sample sizes (Fig. 3e–h). Of course, Maxlike can be used to make similarly robust estimates of relative OP by normalizing its predictions as we did in Fig. 2b.
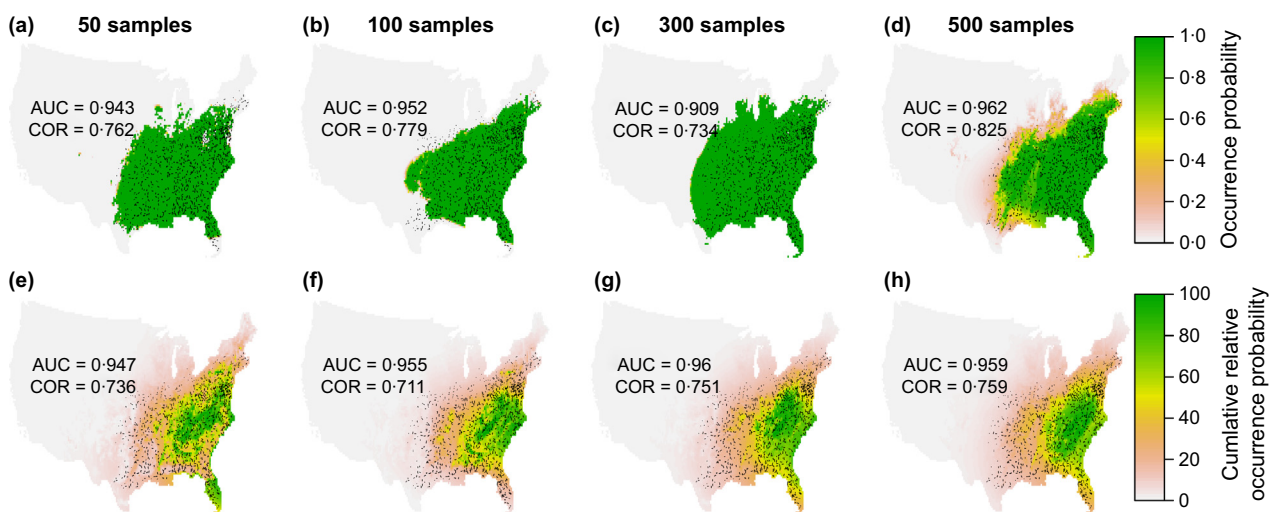

Fig. 3. Comparison of representative Maxlike (a–d) and Maxent (e–h) performance for different sample sizes. AUC and COR are calculated with respect to PA data. The degradation of Maxlike's predictions of OP for smaller data sets is greater than the degradation of Maxent's predictions due to the challenging in estimating prevalence. Maxlike and Maxent offer similar performance, in terms of ranking cells, as shown by AUC and COR. The poor performance of Maxlike in (c) illustrates the challenge of estimating the intercept (cf. Figs 4 and 5); noise in this particular sample lead to a higher intercept estimate than the samples used in (a,b,d).

Using simulated data, we found that Maxlike and Maxent make nearly identical predictions of relative OP (Fig. A3 in Appendix S1; correlation = 0·9999) in spite of the differences in OP reported by Royle *et al.* (2012). Again, Royle *et al.*'s differences derive from using Maxent's logistic format. When the number of presences used for training was reduced below 2000, variance in estimates of the intercept over replicated data sets increased drastically (Fig. 4a). This increase in variance produces a large range of predicted OP, when using sample sizes of 50–1000 (Fig. 4b). Even for 2000 presence samples, predicted OP varies by more than 0·2, on average.

The instability of intercept estimates by Maxlike leads to large average percentage error in predicted OP (Fig. 4c). In contrast, by avoiding estimating the intercept and focusing on relative OP, Maxent's predictions are much less biased (lower % error in Fig. 4d compared with Fig. 4c). While relative OP is less informative than OP, it can be estimated more accurately for smaller sample sizes.

Maxlike struggled to estimate the intercept when the true values of OP did not range from 0 to 1 (Fig. 5). When the slope parameter was smaller, a wider range of values for intercept were possible that were consistent with the slope because OP did not span [0,1] over the range of the covariate. When OP spans [0,1] (Fig. 5a), the logistic curves were much better constrained (i.e. it was not possible to move the logistic curve up/down) and parameter estimates had lower variance.

Discussion

SIMILARITIES BETWEEN MAXLIKE AND MAXENT

It is not surprising that Maxlike and Maxent identify similar environmental relationships given the formal relationship between their slope coefficients. In fact, Manly, McDonald & Thomas (1993) recommend using logistic regression (the form of Maxlike) with the background cells treated as absences as a shortcut to obtaining slope parameters for an exponential model (the form of Maxent) precisely because the slope estimates of one model are well approximated by the other (we observed this in Table 1). Both methods make similar predictions of spatial distributions (Fig. 2), similarity to the GLM using the full PA data set (Fig. 2), predictive performance (AUC in Fig. 2), coefficient values (Table 1) and ranking the predicted suitability of cells (Fig. A2 in Appendix S1). This similarity occurs because the additional information estimated by Maxlike, the intercept in equation 2, primarily slides the predictions up or down the scale of OP (Fig. 5; except to the extent that the intercept and slopes are correlated). Only minor differences are expected in cell rankings due to the different parametric forms of Maxlike and Maxent (eqns 2 and 3) with deviations greatest in the tails of the distributions (Fig. A3 in Appendix S1; Johnson *et al.* 2006). The similarity in slope coefficients means that the differences in predictions between Maxlike and Maxent shown by Royle *et al.* (2012); reproduced in Fig. 2d–f) can be substantially reduced by focusing on

relative OP, as we have done here, or using a different value for the prevalence if such information were known (Fig. A1 in Appendix S1). Note that the coefficients are not expected to be exactly equivalent because different parameter optimization techniques are used by each method.

Many applications of SDMs do not require OP, and relative OP is sufficient, which means that Maxlike's and Maxent's predictions can be interpreted similarly. The ratio of relative OP between cells is equivalent to the ratio of OP, so applications comparing or ranking or ratios of OP in cells are not affected by which method is chosen, for example locating the best habitat for finding a new population or defining priority conservation locations.

PRACTICAL LIMITATIONS OF MAXLIKE

Our explorations with Maxlike suggest that it may exhibit some practical limitations. Our concern is not the fundamental specification of the model, but whether the types of data sets to which it is applied contain sufficient information to provide useful estimates of OP. The parametric assumptions of Maxlike have already been criticized (Hastie & Fithian 2013; Phillips & Elith 2013), who highlight how the parametric (logistic) assumption of Maxlike is fairly restrictive and leads to inaccurate predictions with even mild departures from an exactly loglinear response. While parametric assumptions may be limiting in some cases, very flexible response curves can be readily accommodated using functional transformations of predictors, splines or nonlinear smoothers, so this factor is only partly limiting. Hence, we see the parametric assumptions as only partly limiting and focus on the limitations due to typical data sets and model interpretation in the case where Maxlike's assumptions are acceptable.

The apparent advantage of Maxlike is the ability to estimate the OP, whereas Maxent can only estimate relative OP or ROR. Of course, absolute probabilities are preferable because they contain more information than relative probabilities (Pearce & Ferrier 2000). Reliably estimating the prevalence from PO data requires a large number of observations to avoid large variance in predictions (Figs 3–5), which are usually not available from museum records or opportunistically collected data. While the rankings of cells or their relative OP are reliably estimated by Maxlike for samples of any size, variability in predicted OP can be so high for small sample sizes that the prediction may not be particularly useful (Figs 3 and 4). This highlights the need to report interval estimates for predicted OP. For example, high variance intercept estimates can be problematic when modellers create binary presence–absence predictions because predicted OP can depend heavily on sampling noise (Figs 3 and 4b). Previous studies similarly note the high variability in estimates of the intercept in Maxlike (Fitzpatrick, Gotelli & Ellison 2013) and closely related models (Lancaster & Imbens 1996; Lele & Keim 2006; Ward *et al.* 2009).

Poor predictions made by Maxlike suffer from the difficulty of identifying the intercept when true OP does not span [0,1] (Fig. 5). When true OP does not span [0,1], the response curve can readily shift up or down with little impact on the likelihood
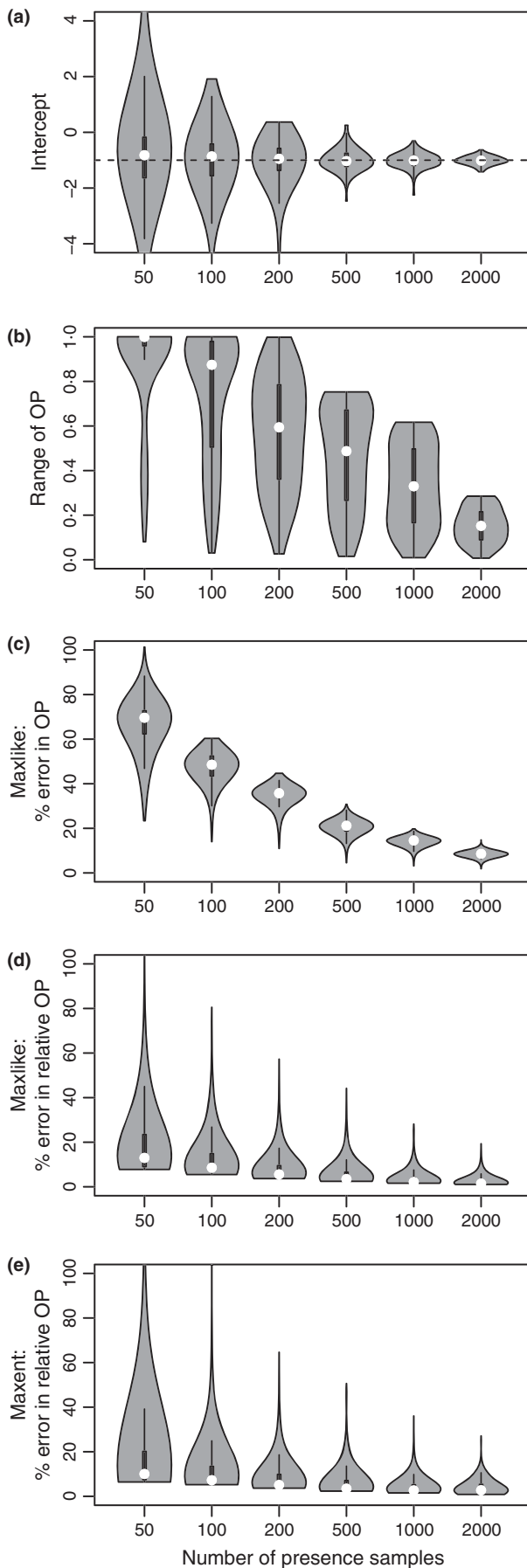
(a)

(b)

(c)

(d)

(e)

Number of presence samples

Fig. 4. Comparison of mean estimates across 500 replicated simulated data sets for different sample sizes using violin plots. White dots indicate the median, black bars indicate quartiles, and grey density functions show the distribution of values. (a) Distribution of the maximum likelihood estimate of the intercept by Maxlike across replicates. The horizontal dashed line at $y = -1$ represents the true value of the intercept. (b) Distribution of range of predicted OP in a cell by Maxlike, averaged over replicated samples and averaged over the 10 000 cells on the simulated landscape. (c) Distribution of average percentage error in predicted OP by Maxlike across replicates. (d) Distribution of average percentage error in predicted relative OP by Maxlike across replicates. (e) Distribution of average percentage error in predicted relative OP by Maxent across replicates. Notably, Maxlike and Maxent have similar predictive precision and bias when estimating relative OP (d–e).

(Fig. 5). This observation also explains Maxlike's poor predictions in Phillips & Elith (2013). In order to apply Maxlike confidently to predict OP, one must determine whether the landscape spans OP values between 0 and 1, whether this is possible in practice *a priori* is an open question.

Substantial challenges exist in estimating sampling effort and detection probability for PO data (Dorazio 2012; Yackulic *et al*. 2012). Thus, the best one can do for typical PO data sets where sampling protocols are unknown is to indicate relative relationships in occurrence, which trivializes the debate over whether one can reliably predict OP vs. relative OP. If independent presence–absence data are available, it is possible to calibrate relative OP predictions to predict OP (Aarts, Fieberg & Matthiopoulos 2012; Halvorsen 2012; Halvorsen 2013). Otherwise, unless a model for detection probability is explicitly included (cf. Kery, Gardner & Monnerat 2010; Chakraborty *et al*. 2011; Yackulic *et al*. 2012), estimates of OP are only defined with respect to a given level of sampling effort (i.e. detection probability and true OP are confounded; Elith *et al*. 2011; Yackulic *et al*. 2012). Thus, interpretation of OP may not be as appealing as it initially appears if detection probability is ignored, as is typically done (Dorazio 2012; Yackulic *et al*. 2012).

Low variance Maxlike predictions may be restricted to sufficiently large data sets that may be well approximated as PA data. This is apparent if we recognize that PO and PA data lie on a continuum defined by sampling effort. If sampling effort is low or unknown, we often do not believe that a lack of presence corresponds to an absence, and we call it PO data. With increasing sampling effort we are more likely to believe these absences and call the observations PA data. This connection is made formally using models for detection probability (e.g. Gelfand *et al*. 2006; Kery, Gardner & Monnerat 2010; Chakraborty *et al*. 2011). In cases where large structured surveys exist, such as the breeding bird survey (from which the wren data derive), presence–absence methods are already a reasonable approximation. It is no surprise that Maxlike converges on similar predictions to a GLM (Fig. 2); the models have the same parametric form (logistic), and the data are nearly identical.

MODELLING DECISIONS WITH MAXENT

Many of the limitations of Maxent arise from choices made during model building, particularly those raised by Royle *et al*.
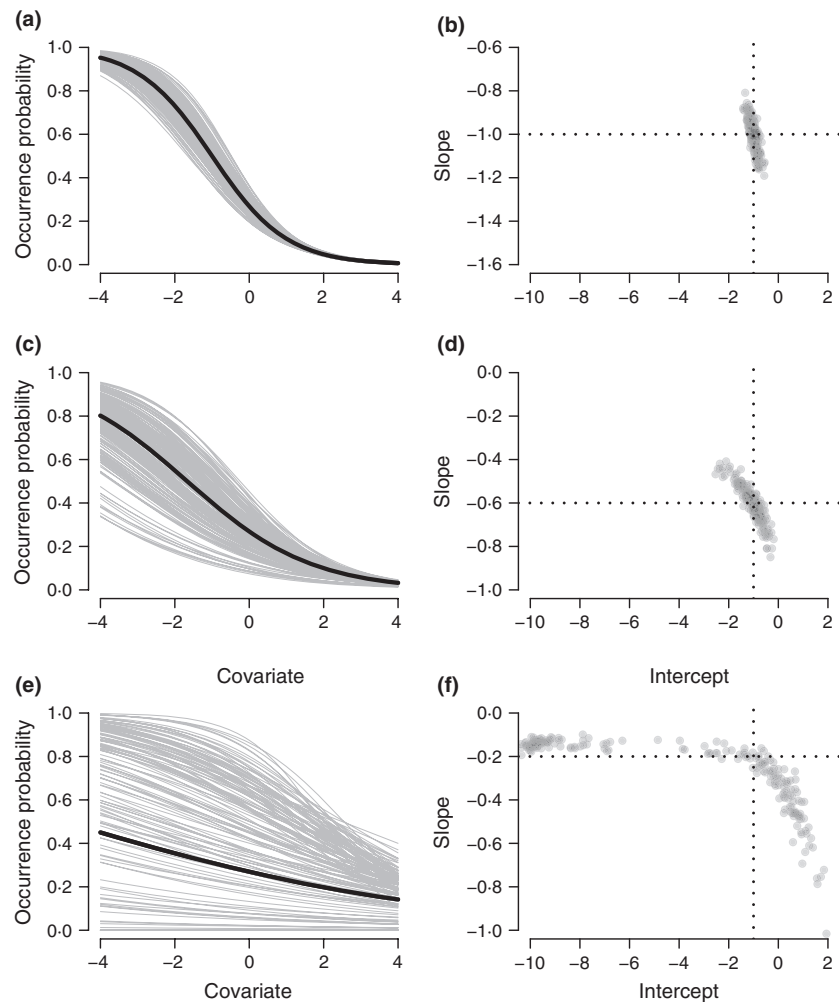
Fig. 5. Maxlike can identify the intercept well when the true OP spans from 0 to 1 because the steep slope parameter does not allow the curve shift vertically (a). In contrast, when the species response is shallower (c, e) the same slope parameter is consistent with a wide range of intercept values (d,f). (a,c,e): Response curves from 200 simulated data sets simulated from (a) $\psi(z) = -1-1*z$, (c) $\psi(z) = -1-0{\cdot}6*z$ and (e) $\psi(z) = -1-0{\cdot}2*z$. (b,d,f) Maximum likelihood parameter estimates from the 200 simulated data sets. Dashed lines indicate true parameter values.

(2012). However, these are not fundamental attributes of Maxent models and are readily adjusted within the software package options (Merow, Smith & Silander 2013), or outside the software package (Halvorsen 2012), although this appears to be rarely done.

Maxent's *logistic output* should not be interpreted as OP because it relies on an assumption, rather than an estimate, of prevalence (Elith *et al.* 2011; Royle *et al.* 2012; Merow, Smith & Silander 2013). The differences between Maxlike and Maxent in Royle *et al.* (2012) emerged because they used Maxent's logistic output to predict OP; they have highlighted why this is problematic. Figure A1 in Appendix S1 shows the high variability in predictions that can emerge for different prevalence assumptions for the Carolina wren. However, these concerns about logistic output do not affect the interpretation of Maxent's raw output as relative OP; hence, we recommend that Maxent analyses focus on relative OP or ROR.

Maxent builds very flexible occurrence–environment relationships, which have been criticized for overfitting (Warren & Seifert 2011; Royle *et al.* 2012; Merow, Smith & Silander 2013). Maxent uses a number of functional transformations of each predictor, known as features, in conjunction with regularization to penalize for model complexity and select the best features from a very large candidate set (Tibshirani 1996;

Phillips, Anderson & Schapire 2006; Merow, Smith & Silander 2013). Explorations with complex models can be valuable to explore whether any unexpected patterns exist in response curves but can be dangerous to interpret because such flexibility is bound to extract some pattern even when no causal relationships exist. One approach to reducing complexity in Maxent is to simply reduce the number of feature classes that it considers (e.g. only linear and quadratic features). Maxent's software package does not offer the option of including just a few specific features (e.g. based on specific hypotheses) but instead takes an all-or-nothing approach, which makes it difficult to evaluate the importance of specific features. If a subset of more complex features are desired, one can ensure that the model is built with precisely the user-specified set of features by (i) making the features manually, (ii) providing them as a layers to Maxent, (iii) selecting only linear features and (iv) optionally setting the regularization multiplier to zero (Merow, Smith & Silander 2013). Stronger control of the features included and model selection is also possible using an IPP framework (Renner & Warton 2013) or using insights from the strict maximum likelihood interpretation of Maxent (Halvorsen 2013).

The use of regularization has been criticized (Royle *et al.* 2012), although it is a widely accepted tool in both machine

learning and statistics (Tibshirani 1996; Hastie, Tibshirani & Friedman 2009). If regularization is undesirable, Maxent's regularization parameter can easily be set to zero to maximize the unpenalized likelihood. The appeal of regularization is that it reduces overfitting to bias in the data. Choosing the appropriate strength of the regularization penalty is a more difficult matter that has received little attention (but see Phillips & Dudik 2008; Elith, Kearney & Phillips 2010; Anderson & Gonzalez 2011; Merow, Smith & Silander 2013; Renner & Warton 2013).

RECOMMENDATIONS

The connection between Maxent and Maxlike leaves users with a conundrum for many practical applications. When using Maxent as an OP model, one can make a good assumption about PO data – that it represents a random sample of space – but is forced to predict only relative OP due to limitations of the loglinear model. Alternatively, when using Maxent as an ROR model, one can make a questionable assumption about PO data – that it represents a random sample of individuals – and make a good assumption that the loglinear model is appropriate for count data. Maxlike makes reasonable assumptions about both the data and model for predicting OP. However, there exist a limited set of circumstances where the predicted OP has sufficiently low variance to be more informative than relative OP, due to the limited information contained in PO data.

Modelling relative OP is valuable whenever interpretations do not rely on absolute values of OP. We expect that this scenario is rather common, given the exploratory nature of many occurrence models. This recommendation derives from the challenges of interpreting (absolute) OP, due to unknown detection probability and sampling effort (Section Practical limitations of Maxlike; Yackulic *et al.* 2012). Modelling relative OP can be done with either Maxent or Maxlike, though Maxent may be preferable if correlation between the intercept and slope parameters is present (Fig. 5). If the number of samples is large, OP spans [0,1], and detection probability is constant, Maxlike is preferable because it can predict OP. In any case, the prediction uncertainty should always be reported given its large magnitude in the relatively well-behaved case studies shown here (Figs 3–5).

When data represent a random sample of individuals, we recommend using Maxent to model ROR, or more generally a model within the IPP family (cf. Aarts, Fieberg & Matthiopoulos 2012; Fithian & Hastie 2013). Although Maxent is suitable when the landscape must be discretized and its settings have been carefully selected (cf. Merow, Smith & Silander 2013), working in an IPP framework offers a few advantages over Maxent. IPPs can provide explicit connections between OP and ROR (eqn 9; Baddeley *et al.* 2010), and estimate the intercept of a loglinear model (Aarts, Fieberg & Matthiopoulos 2012) when total abundance is measured in at least some cells. IPPs can be fit using a range of software for all flavours of generalized linear models, for example GAMs, boosted regression trees or ridge/penalized regression, that come with a variety of

tools for model assessment (Fithian & Hastie 2013). If information is available on abundance, the intercept parameter of the IPP can be estimated once the slope parameters have been estimated describing to ROR (see Appendix from Aarts, Fieberg & Matthiopoulos 2012). Methods exist to determine the optimal spatial resolution and number of background points from the gridded landscape for an IPP, rather than the *ad hoc* methods typically used with Maxent (Warton & Shepherd 2010; Renner & Warton 2013). While OP from models built at different spatial resolutions typically cannot be compared, IPPs facilitate this comparison using intensity functions and offset terms based on pixel size during model fitting (Baddeley *et al.* 2010). Renner & Warton (2013) provide additional advantages, including theory regarding regularization, methods for assessing goodness-of-fit, and discussion of the loss of information associated with the gridded landscapes used in Maxent.

Though the merits of Maxlike and Maxent have been sharply contrasted against one another (Royle *et al.* 2012; Phillips & Elith 2013), we find that the limitations of PO data constrain modellers to focus on relative OP and that Maxlike and Maxent are similarly valuable for predicting relative OP (Fig. 4d, e) once modelling decisions have been carefully made. Consequently, we advocate for informed usage of any species distribution modelling method, by focusing on understanding assumptions, hypotheses and interpretations.

Acknowledgements

References

Aarts, G., Fieberg, J. & Matthiopoulos, J. (2012) Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution*, **3**, 177–187.

Anderson, R.P. (2012) Harnessing the world's biodiversity data: promise and peril in ecological niche modelling of species distributions. *Annals of the New York Academy of Sciences*, **1260**, 66–80.

Anderson, R.P. & Gonzalez, I. (2011) Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent. *Ecological Modelling*, **222**, 2796–2811.

Araújo, M.B. & Peterson, A.T. (2012) Uses and misuses of bioclimatic envelope modelling. *Ecology*, **93**, 1527–1539.

Baddeley, A., Berman, M., Fisher, N.I., Hardegen, A., Milne, R.K., Schuhmacher, D., Shah, R. & Turner, R. (2010) Spatial logistic regression and change-of-support in Poisson point processes. *Electronic Journal of Statistics*, **4**, 1151–1201.

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. & Silander, J.A. (2011) Point pattern modelling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society: Series C (Applied statistics)*, **60**, 757–776.

Della Pietra, S., Della Pietra, V. & Lafferty, J. (1997) Inducing features of random fields. *IEEE Transactions: Pattern Analysis and Machine Intelligence*, **19**, 1–13.

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

Dudik, M., Phillips, S. & Schapire, R. (2004) Performance guarantees for regularized maximum entropy density estimation. *Learning Theory, Proceedings*, **3120**, 472–486.

Elith, J., Kearney, M. & Phillips, S. (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution*, **1**, 330–342.

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

Fielding, A. & Bell, J. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

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

Fitzpatrick, M.C., Gotelli, N.J. & Ellison, A.M. (2013) MaxEnt versus MaxLike: empirical comparisons with ant species distributions. *Ecosphere*, **4**, art 55.

Gelfand, A.E. *et al.* (2006) Explaining species distribution patterns through hierarchical modelling. *Bayesian Analysis*, **1**, 42–92.

Graham, C. *et al.* (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.

Halvorsen, R. (2012) A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. *Sommerfeltia*, **36**, 1–132.

Halvorsen, R. (2013) A gradient analytic perspective on distribution modelling. *Sommerfeltia*, **35**, 1–165.

Hastie, T. & Fithian, W. (2013) Inference from presence-only data; the ongoing controversy. *Ecography*, **36**, 1–4.

Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn. Springer-Verlag, New York.

Johnson, C.J. *et al.* (2006) Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. *The Journal of Wildlife Management*, **70**, 347–357.

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

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P. & Jetz, W. (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in Ecology and Evolution*, **4**, 82–94.

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

Lancaster, T. & Imbens, G. (1996) Case-control studies with contaminated controls. *Journal of Econometrics*, **71**, 145–160.

Lele, S.R. (2009) A new method for estimation of resource selection probability function. *The Journal of Wildlife Management*, **73**, 122–127.

Lele, S.R. & Keim, J.L. (2006) Weighted distributions and estimation of resource selection probability functions. *Ecology*, **87**, 3021–3028.

Manly, B.F.J., McDonald, L.L. & Thomas, D.L. (1993) *Resource Selection by Animals: Statistical Analysis and Design for Field Studies*, 1st edn. Chapman & Hall, London.

Merow, C., Smith, M.J. & Silander, J.A. (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1–12.

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

Phillips, S., Anderson, R. & Schapire, R. (2006) Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

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

Phillips, S.J. & Elith, J. (2013) On estimating probability of presence from use-availability or presence-background data. *Ecology*, **94**, 1409–1419.

Phillips, S.J. *et al.* (2004) A maximum entropy approach to species distribution modelling. *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 472–486. ACM Press, New York.

Phillips, S. *et al.* (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**, 181–197.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

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

Royle, J.A. *et al.* (2012) Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. *Methods in Ecology and Evolution*, **3**, 545–554.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, **58**, 267–288.

Ward, G. *et al.* (2009) Presence-only data and the EM algorithm. *Biometrics*, **65**, 554–563.

Warren, D. & Seifert, S. (2011) Ecological niche modelling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, **21**, 335–342.

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

Yackulic, C.B., Chandler, R., Zipkin, E.F., Royle, J.A., Nichols, J.D., Campbell Grant, E.H. & Veran, S. (2012) Presence-only modelling using MAXENT: when can we trust the inferences? *Methods in Ecology and Evolution*, **4**, 236–243.

Zheng, B. & Agresti, A. (2000) Summarizing the predictive power of a generalized linear model. *Statistics in Medicine*, **19**, 1771–1781.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Additional examples.

Figure A1. Comparison of Maxent predictions for different values of Maxent's τ parameter.

Figure A2. (a) Comparison of cell rankings of 12 082 points (with multiple records occurring within a cell) from the full PO data set (used to build the GLM) between Maxlike and Maxent for the Carolina wren model shown in Fig. 2. (b) Comparison of predicted cumulative values for the same PA data as (a). There is a larger disparity in the ranking of low suitability cells, deriving from Maxent's more spatially diffuse predictions.

Figure A3. Comparison of relative OP predictions Maxlike and Maxent for simulated data.

Data S1. R script files 'maxlike simulated data'.

Data S2. R script files maxlike wren data.