

# Equivalence of MAXENT and Poisson Point Process Models for Species Distribution Modeling in Ecology

Ian W. Renner\* and David I. Warton

School of Mathematics and Statistics and Evolution & Ecology Research Centre,  
The University of New South Wales, NSW 2052, Australia.

\*email: Ian.Renner@unsw.edu.au

**SUMMARY.** Modeling the spatial distribution of a species is a fundamental problem in ecology. A number of modeling methods have been developed, an extremely popular one being MAXENT, a maximum entropy modeling approach. In this article, we show that MAXENT is equivalent to a Poisson regression model and hence is related to a Poisson point process model, differing only in the intercept term, which is scale-dependent in MAXENT. We illustrate a number of improvements to MAXENT that follow from these relations. In particular, a point process model approach facilitates methods for choosing the appropriate spatial resolution, assessing model adequacy, and choosing the LASSO penalty parameter, all currently unavailable to MAXENT. The equivalence result represents a significant step in the unification of the species distribution modeling literature.

**KEY WORDS:** Habitat modeling; Location-only; Maximum entropy; Poisson likelihood; Presence-only data; Use-availability.

## 1. Introduction and Background

Species distribution modeling (SDM), where the goal is to explain the occurrence of a species using a set of environmental variables, is an important goal in ecology. This is a fast-moving field; in fact ISI's Essential Science Indicators (July 2012) identifies SDM as one of the top five ranked research fronts in ecology and the environmental sciences. One potential reason for such high interest is that SDM aims to address important topical questions such as the potential effects of climate change on species distributions (Thullier et al., 2008). Rapid progress in this field has been facilitated by recent significant technological advances in remote sensing, GIS (O'Sullivan and Unwin, 2010), and computational power, enabling models to be built at increasingly fine resolutions and increasingly large spatial scales.

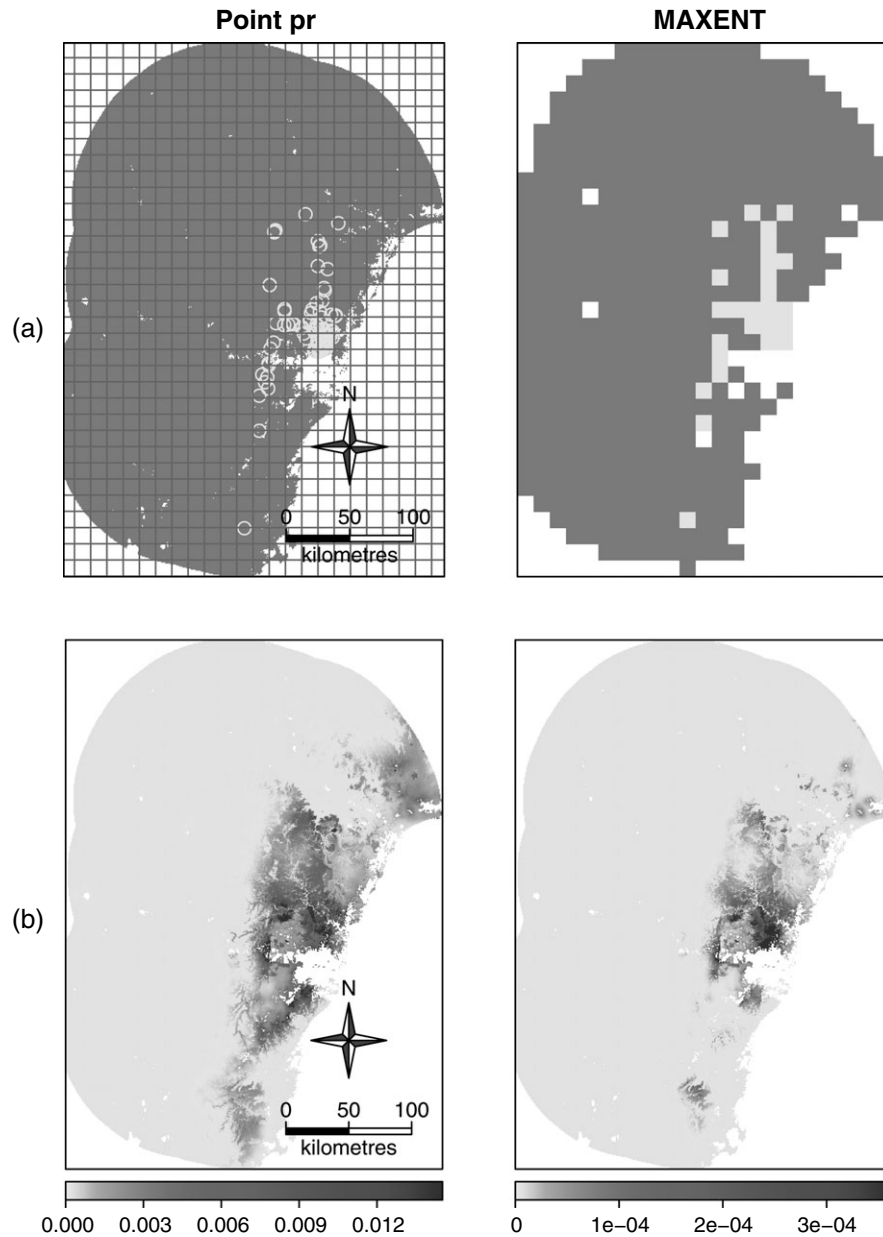
Ideally, an SDM could be constructed using systematically collected presence/absence data so that logistic regression (McCullagh and Nelder, 1989) and its extensions (Hastie and Tibshirani, 1990; Elith, Leathwick, and Hastie, 2008) may be used. However, the best available data often come not from systematic data but from lists of locations where a species is reported to be present, with no corresponding information about where a species is reported to be absent (Pearce and Boyce, 2006). This type of data, known as "presence-only" data, is typically found in museums, herbaria, and atlases (Pearce and Boyce, 2006).

An example used throughout this article is a list of 95 locations (NSW Office of Environment and Heritage, 2010) of Sydney eucalypt *Corymbia eximia* observed between 1990 and 2008 within 100 km of the Greater Blue Mountains World

Heritage Area, Australia (Figure 1a). We would like to model the distribution of *C. eximia* as a function of climatic and fire history variables in order to explore the nature of association of each of these variables with occurrence of this species.

MAXENT (Phillips, Anderson, and Schapire, 2006), based on a maximum entropy approach, is particularly common in SDM, having been cited 378 times in 2011 according to Google Scholar. Its rise in popularity has been meteoric, having only been introduced to ecology 6 years ago, although the concept of maximum entropy modeling has been around for a long time (Jaynes, 1957). A comprehensive study of current SDM methods found MAXENT to outperform nearly all other methods (Elith et al., 2006), and this may explain its prevalence in the literature. Nevertheless, MAXENT has a number of shortcomings, as demonstrated in Sections 3 and 4. In particular, it is unclear what diagnostic tools may be used to assess whether the fitted model is reasonable. Moreover, MAXENT analyzes data after first aggregating them into presence/absence grid cells (as in Figure 1a), and it is currently unclear what spatial resolution should be used when constructing these grid cells. Further, some key components of the output such as the intercept and fitted probabilities are dependent on this choice of spatial resolution ("scale-dependence").

In this article we show that MAXENT is mathematically equivalent to Poisson regression (McCullagh and Nelder, 1989) and related to a Poisson point process model (Warton and Shepherd, 2010). Relationships between maximum likelihood and maximum entropy have been known for a long



**Figure 1.** Comparison of point process and MAXENT analyses of *Corymbia eximia* data. (a) Response variables for *C. eximia*: a Poisson point process model (left) or an area-interaction point process model (left) analyzes presence points  $y_p = \{y_1; \dots; y_m\}$  MAXENT (right) analyzes presence/absence in grid cells  $\{g_1, \dots, g_n\}$ , with  $n = 258$  here. A key issue with MAXENT is determining how many grid cells  $n$  to use for analysis. (b) Predicted species distribution maps for an area-interaction model (left) and MAXENT (right). This figure appears in color in the electronic version of this article.

time—this relationship was explored for exponential families in the late 1950s (Kullback, 1959), while an equivalence for contingency tables was established in 1963 (Good, 1963), and maximum entropy was later linked to the maximum likelihood of a Gibbs distribution (Della Pietra, Della Pietra, and Lafferty, 1997). Nonetheless, the direct link we make in this article between MAXENT and Poisson point process models is new. Warton and Shepherd (2010) introduced Poisson point process models as a way to address “problems of model specification, interpretation, and implementation” inherent in

pseudo-absence regression, another popular method of SDM. This article achieves a similar goal in relation to MAXENT—all of the problems described in Sections 3 and 4 can be addressed by reframing the problem using a Poisson point process model. Section 2 demonstrates the equivalence of Poisson point process models and MAXENT. Section 3 demonstrates by example how this equivalence can improve on current practice in MAXENT modeling. Finally, Section 4 demonstrates that these proposed improvements can lead to more accurate predictions of a species’ actual distribution.

**2. Equivalence of MAXENT and Poisson Point Process Models**

The goal of SDM is to link the location of species presences to some number  $p$  of environmental variables. Let  $\mathbf{y}_P = \{y_1, \dots, y_m\}$  be presence-only locations for a particular species over some region  $\mathcal{A}$  and  $\mathbf{x}(y)' = \{1, x_1(y), \dots, x_p(y)\}'$  be the vector of  $p$  environmental variables corresponding to location  $y$  in the study region  $\mathcal{A}$ . We fit an SDM by regressing the  $y \in \mathcal{A}$  against the  $\mathbf{x}(y)$ , using one of a few methods.

Rather than using the presence-only locations  $\mathbf{y}_P = \{y_1, \dots, y_m\}$ , the MAXENT procedure analyzes data by splitting  $\mathcal{A}$  into  $n$  grid cells with centers at the locations in  $\mathbf{g} = \{g_1, \dots, g_n\}$ . A binary response vector  $\mathbf{z}^{(n)}(\mathbf{g}) = \{z^{(n)}(g_1), \dots, z^{(n)}(g_n)\}$  is formed where  $z^{(n)}(g_i) = 1/m^{(n)}$  if the  $i$ th grid cell contains at least one presence location and 0 otherwise, and  $m^{(n)}$  is the count of grid cells that contain at least one presence location. Without loss of generality, we partition  $\{g_1, \dots, g_n\}$  as  $\{\mathbf{g}_P, \mathbf{g}_0\}$ , where  $\mathbf{g}_P = \{g_1, \dots, g_{m^{(n)}}\}$  are the  $m^{(n)}$  presence cells. We index  $z$  and  $m$  with the superscript  $(n)$  to emphasize that these quantities depend on the spatial resolution used for resolution (hence the number of grid cells  $n$ ) used in analysis. The goal in MAXENT is to model  $\pi(g_i)$ , the probability that if there is one presence then it is located in the  $i$ th grid cell.  $\boldsymbol{\pi}(\mathbf{g}) = \{\pi(g_1), \dots, \pi(g_n)\}$  is estimated to maximize the entropy  $H\{\boldsymbol{\pi}(\mathbf{g})\} = -\sum_{i=1}^n \pi(g_i) \ln \pi(g_i)$ , subject to two types of constraint:

$$\sum_{i=1}^n \pi(g_i) x_j(g_i) = \frac{1}{m^{(n)}} \sum_{i=1}^{m^{(n)}} x_j(g_i), \quad \forall j, \quad (1)$$

$$\sum_{i=1}^n \pi(g_i) = 1. \quad (2)$$

Equation (1) ensures that the predicted mean of each environmental variable equals its observed mean for the presence data while (2) ensures that the probabilities add to 1.

We will show that the MAXENT procedure is equivalent to Poisson regression when applied to grid cell data  $\mathbf{z}^{(n)}(\mathbf{g})$ . That is, we model the mean of  $z^{(n)}(g_i)$  as a log-linear model:

$$\ln \mu_i = \mathbf{x}(g_i)' \boldsymbol{\beta}. \quad (3)$$

We estimate the parameters  $\boldsymbol{\beta}$  to maximize the likelihood function (McCullagh and Nelder, 1989):

$$l\{\boldsymbol{\beta}; \mathbf{z}^{(n)}(\mathbf{g})\} = \sum_{i=1}^n z^{(n)}(g_i) \ln \mu(g_i) - \sum_{i=1}^n \mu(g_i) - \sum_{i=1}^n \ln\{z^{(n)}(g_i)!\}. \quad (4)$$

On face value, this analysis appears to be based on a nonsensical model for the data, as it implicitly assumes that a set of noninteger values comes from a Poisson distribution. However, we will show first that this is precisely what MAXENT does and later that this can be motivated as a point process model, which can be fitted for a noninteger response using the result of Berman and Turner (1992).

**THEOREM 1.** *The MAXENT procedure and Poisson regression are equivalent. That is,*

1. They fit the same model:

$$\ln \pi(g_i) = \ln \mu(g_i) = \mathbf{x}(g_i)' \boldsymbol{\beta}.$$

2. They estimate parameters to maximize the same function up to a constant:

$$\Lambda\{\boldsymbol{\beta}; \mathbf{z}^{(n)}(\mathbf{g})\} = l\{\boldsymbol{\beta}; \mathbf{z}^{(n)}(\mathbf{g})\} + C,$$

where  $C$  is a constant and  $\Lambda\{\boldsymbol{\beta}; \mathbf{z}^{(n)}(\mathbf{g})\}$  is the Lagrangian function to maximize entropy  $H\{\boldsymbol{\pi}(\mathbf{g})\}$  subject to the constraints stated in equations (1)–(2). Hence the maximum entropy estimate  $\hat{\boldsymbol{\beta}}_{\text{MAXENT}}$  equals the maximum likelihood estimate from Poisson regression  $\hat{\boldsymbol{\beta}}_{\text{GLM}}$ .

The proof of Theorem 1 appears in Web Appendix 1. Part 1 of Theorem 1 (that MAXENT fits a log-linear model) is well known (e.g. Dutta, 1966), but Part 2 (the link to Poisson regression) is new. This link to Poisson likelihood was enabled by specifying the MAXENT model in a slightly different way to what is conventional in the maximum entropy literature. It is typical to exclude the intercept term from the model and introduce a normalization constant in its place after optimization to ensure that the sum of  $\boldsymbol{\pi}$  is one. Instead, we included an intercept term and the constraint of equation (2) to the optimization problem, which was the key to our derivation. Hence we have shown that some maximum entropy problems, including MAXENT, can be solved using *standard generalized linear modeling software* via Poisson regression, which can accommodate a large number of predictors. We demonstrate this result numerically in Web Figure 1. Further, this enables a link with Poisson point process models below.

A Poisson point process regression model (PPM) analyzes  $m$  presence-only locations  $\mathbf{y}_P = \{y_1, \dots, y_m\}$  as a point process in which the locations of the  $m$  points are assumed to be independent. Unlike MAXENT, which models probability  $\pi(g_i)$  per grid cell, a Poisson PPM models the limiting expected count ( $\lambda(y)$ , the “intensity”) *per unit area* (Cressie, 1993) for any location  $y \in \mathcal{A}$ . Intensity is modeled as a log-linear function of  $p$  explanatory variables:  $\ln\{\lambda(y)\} = \mathbf{x}(y)' \boldsymbol{\beta}$ . An analysis on a per area basis rather than a per grid cell basis is a key distinction between a point process model and MAXENT.

The log-likelihood of a Poisson point process model (Cressie, 1993) is:

$$l(\boldsymbol{\beta}; \mathbf{y}_P) = \sum_{i=1}^m \ln \lambda(y_i) - \int_{y \in \mathcal{A}} \lambda(y) dy - \ln(m!). \quad (5)$$

By using numerical quadrature (Davis and Rabinowitz, 1984), the likelihood expression for  $\mathbf{y}_P$  can be approximated as a weighted Poisson likelihood (Berman and Turner, 1992):

$$l_{\text{ppm}}(\boldsymbol{\beta}; \mathbf{y}_P, \mathbf{y}_0, \mathbf{w}) \approx \sum_{i=1}^n w_i [z_{w,i} \ln\{\lambda(y_i)\} - \lambda(y_i)], \quad (6)$$

where  $\mathbf{y}_0 = \{y_{m+1}, \dots, y_n\}$  are quadrature points and  $z_{w,i} = \frac{I(i \in \{1, \dots, m\})}{w_i}$  for quadrature weights  $\mathbf{w} = \{w_1, \dots, w_n\}$ , and  $I(\cdot)$  is the indicator function. A natural way to choose quadrature points is to break the region  $\mathcal{A}$  into a regular grid and insert

a quadrature point at the center of each cell, meaning that  $\mathbf{y}_0 = \mathbf{g}_0$ . Each cell can then be assigned a quadrature weight which equals its area divided by the number of locations in  $\{\mathbf{y}_P, \mathbf{y}_0\}$  contained in the cell.

An alternative representation of the point process likelihood, suggested during review, was to use  $I(i \in 1, \dots, m)$  as the response and  $\ln w_i$  as an offset term. This would produce a likelihood expression proportional to (5), but without the need for a noninteger response.

We find a relation in Theorem 2 between MAXENT and the above formulation for Poisson point process models by analyzing data at grid cell locations  $\{\mathbf{g}_P, \mathbf{g}_0\}$  instead of  $\{\mathbf{y}_P, \mathbf{y}_0\}$ . That is, we use in the analysis the same quadrature points  $\mathbf{y}_0 = \mathbf{g}_0$ , but use the locations of the  $m^{(n)}$  presence grid cells  $\mathbf{g}_P$  in place of the  $m$  actual presence locations in  $\mathbf{y}_P$ . This results in some loss of information, discussed in Section 3.

**THEOREM 2.** *Consider a Poisson point process model fitted to grid cell data  $\mathbf{z}^{(n)}(\mathbf{g})$ , with parameter estimates stored in  $\hat{\beta}_{\text{PPM}}$ . Then:*

$$\hat{\beta}_{\text{MAXENT}} = \hat{\beta}_{\text{PPM}} + J_C,$$

where  $J_C = \{\ln C, 0, \dots, 0\}$  is a vector of length  $p + 1$ , and  $C = |\mathcal{A}|/m^{(n)}n$ .

In other words, the MAXENT and PPM solutions for grid cell data are proportional, and estimates of slope parameters are identical.

The proof of Theorem 2 appears in Web Appendix 1.

**COROLLARY 1:** *For a given presence-only dataset  $\mathbf{y}_P$ , consider a set of vectors of grid cell data constructed at increasingly fine spatial resolutions (e.g., by recursively partitioning  $\{\mathbf{z}^{(n)}(\mathbf{g}); n = 1, 2, 2^2, 2^3, \dots\}$ ). As  $n \rightarrow \infty$ , the MAXENT solution for  $\mathbf{z}^{(n)}(\mathbf{g})$  becomes proportional to the Poisson point process model solution for  $\mathbf{y}_P$ . That is:*

$$\hat{\beta}_{\text{MAXENT}} - J_C \rightarrow \hat{\beta},$$

where  $J_C$  is as defined in Theorem 2.

The proof follows by noting that as  $n \rightarrow \infty$ , the number and location of presence points in  $\mathbf{g}_P$  approach those in  $\mathbf{y}_P$  and the quadrature approximation of (6) approaches the exact solution in (5).

This result is similar to Theorem 3.2 of Warton and Shepherd (2010), who showed that when fitting a Poisson PPM with constant quadrature weights  $C$ , ignoring these weights changes the solution by the factor  $C$ . MAXENT can be represented as a Poisson point process model ignoring quadrature weights, so a similar result applies. These quadrature weights are the mechanism that ensures that analysis is performed on an area basis instead of a grid cell basis (Warton and Shepherd, 2010). Hence while Poisson point process model and MAXENT solutions are qualitatively identical, analyzing data on a grid cell basis instead of an area basis induces scale dependence in MAXENT: as  $n \rightarrow \infty$ ,  $\pi(g_i) \rightarrow 0$ . Hence the maps in Web Figure 2 look the same, but only for the Poisson point process models is the scale unchanged by changing spatial resolution.

### 3. Model Application

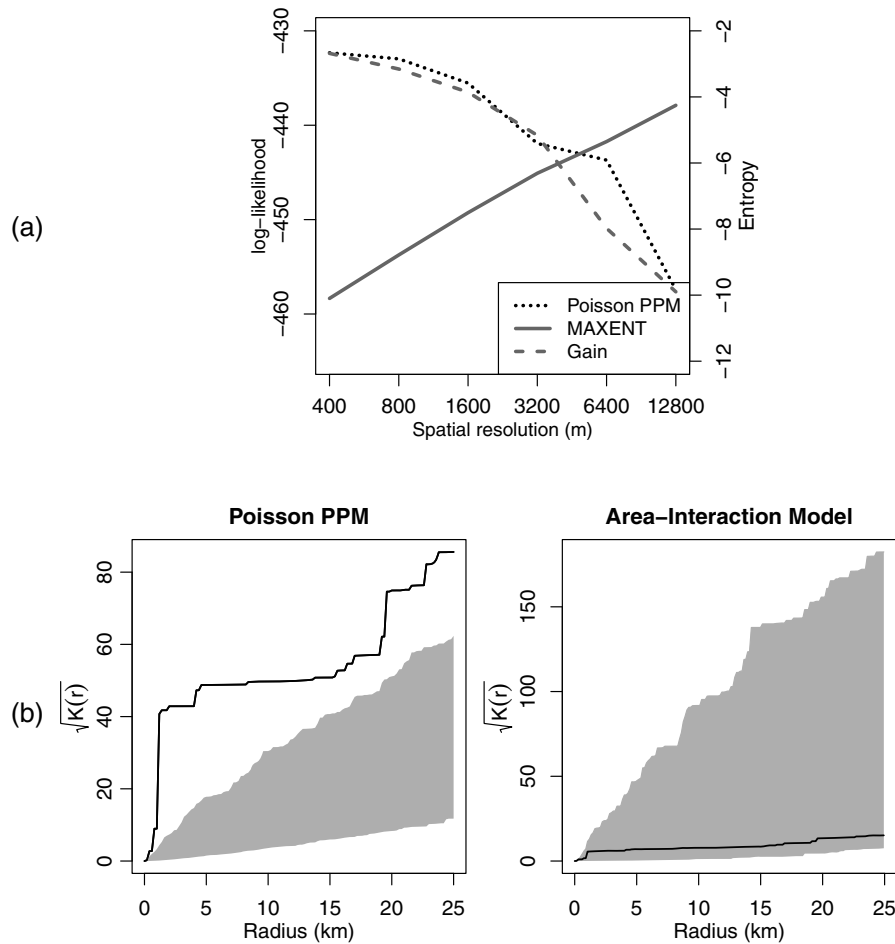
We will now demonstrate the application of a point process model to the presence-only locations of *Corymbia eximia*, illustrating many features currently unavailable to MAXENT. Software for the below analyses including example data will be available in the R package `ppmlasso`. Our analysis will consist of four steps: (1) determine the appropriate spatial resolution for analysis; (2) assess whether a Poisson point process model is appropriate; (3) estimate the LASSO parameter (Tibshirani, 1996) for regularization; and (4) compare results with a MAXENT model. We use four environmental variables as in Warton and Shepherd (2010)—minimum and maximum temperature, number of fires since 1943, and annual rainfall. Likelihood of observing a presence point depends not just on the spatial distribution of the species, but also on the spatial distribution of observers, which is strongly affected by site accessibility. Hence we include two variables to measure site accessibility—distance from main roads and distance from urban areas. Intensity of *C. eximia* was modeled as a quadratic function of the six available variables, including interactions between the four environmental variables and between the two accessibility variables (but assuming additivity between environmental and accessibility variables). So long as all six of these variables are independent of variables associated with species detection probability, parameter estimates from a Poisson point process model will be consistently estimated (Dorazio, in press).

Prior to applying the LASSO to point process models, variables were standardized to have mean 0 and variance 1 as in Tibshirani (1996), such that the LASSO penalty was applied to standardized coefficients. In MAXENT, variables were instead standardized to have minimum 0 and maximum 1.

#### 3.1 Choosing the Appropriate Spatial Resolution

NSW Office of Environment and Heritage (2010) provides environmental data over the study region at the 100 m resolution. However, performing an analysis at such a fine resolution is computationally expensive and may not be necessary. Using a Poisson point process model specification facilitates the use of a numerical integration framework for choosing an appropriate spatial resolution for a particular species. As the absence grid cells  $\mathbf{g}_0$  are used as quadrature points, the question of what spatial resolution needs to be used can be rephrased as a question of how many quadrature points are needed to obtain a sufficiently accurate estimate of the log-likelihood. The same idea was used in Warton and Shepherd (2010) to clarify the role of pseudo-absences in presence-only analysis, and how their number and location can be chosen.

Following Warton and Shepherd (2010), we add quadrature points at increasingly fine resolutions until the log-likelihood has converged. For *Corymbia eximia*, the likelihood appears to converge at a spatial resolution of 800 m (Figure 2a), suggesting that model output will not appreciably change at finer spatial resolutions. However, the entropy of analogous MAXENT models does not converge due to the scale dependence of  $\pi(\mathbf{g})$  and hence MAXENT is not very informative about which spatial resolution to use for analysis. The scale dependence of MAXENT can be adjusted for in part (using “gain,” defined as  $\ln n$  - entropy), but not completely, since the loss of information incurred by absorbing the  $m$  presence



**Figure 2.** Model checking for the *Corymbia eximia* analysis: (a) Spatial resolution can be chosen for a point process model from a plot of maximized log-likelihood at differing spatial resolutions. Convergence is achieved at the 800m resolution for the Poisson point process model, suggesting this is the optimal spatial resolution at which to perform analysis. There is no convergence for the entropy used by MAXENT. We can attempt to address this by analyzing “gain” (defined as  $\ln n$  - entropy), but gain (rescaled) does not converge until the number of presence cells  $m^{(n)}$  converges. (b) Inhomogeneous  $K$ -function (solid line), with 95% simulation envelope (shaded area), for a Poisson point process model (left) and an area-interaction model with radius 1 km (right). The deviation from the envelope for the Poisson point process model suggests additional clustering unaccounted for in the model. This figure appears in color in the electronic version of this article.

locations into a smaller number  $m^{(n)}$  of presence grid cells varies with the choice of spatial resolution. Hence the gain will not converge until  $m^{(n)}$  converges.

### 3.2 Is a Poisson PPM Appropriate?

The underlying assumption of a Poisson point process model (and by equivalence, MAXENT) is that the point locations are independent, conditional on model covariates. This may not be appropriate for *Corymbia eximia*. While MAXENT offers no method for checking this assumption, there are a number of diagnostic tools to assess model adequacy of a Poisson point process model (Cressie, 1993; Baddeley et al., 2005). One such method is to construct the inhomogeneous  $K$ -function (Baddeley, Møller, and Waagepetersen, 2000) and corresponding simulation envelope (Diggle, 2003) of the fitted model. In Figure 2b, it can be seen that for *C. eximia*, a Poisson point process model may not be suitable for the data, as the inhomogeneous  $K$ -function falls well

outside a 95% envelope formed by simulating 1000 realizations from a Poisson point process model with intensity function as estimated from the *C. eximia* data. The deviation above the envelope suggests that the presence locations of *C. eximia* are more clustered than would be expected for a true Poisson point process model. Instead, Figure 2b demonstrates that an area-interaction model (Baddeley and van Lieshout, 1995) with radius 1 km is more appropriate, which we fit using a Poisson pseudo-likelihood as in the `spatstat` (Baddeley and Turner, 2005) package of R. There is built-in code in `spatstat` for fitting a large suite of other spatial processes involving dependence between points (Baddeley and Turner, 2005; Chakraborty et al., 2011) that may be suitable.

### 3.3 Choosing the LASSO parameter

MAXENT is often fitted using a LASSO penalty to control for overfitting. For *Corymbia eximia*, MAXENT software uses an ad hoc value of  $\frac{9}{70}$  for the LASSO penalty parameter ( $\lambda$ ),

**Table 1**

Current problems with MAXENT and their proposed solutions available through reexpression as a Poisson point process model

| MAXENT problem                              | Poisson PPM solution                         |
|---|--|
| Predicted probabilities are scale-dependent | Predicted intensities are scale-invariant    |
| How to determine spatial resolution?        | Increase until log-likelihood converges      |
| How to assess model adequacy?               | Various goodness-of-fit procedures available |
| How to choose LASSO parameter?              | Various data-driven methods                  |
| Available in MAXENT software only           | Use any standard GLM software                |
| 130 seconds to fit models in Figure 1b      | 12 seconds to fit models in Figure 1b        |

which was chosen without any consideration for predictive performance of the model at hand but rather based entirely on the number of presence cells (90). Alternatively, some data-driven criterion could be used to try to choose a  $\lambda$  which optimizes predictive performance (Tibshirani, 1996; Fu, 2005; Zou, Hastie, and Tibshirani, 2007). We used a simple line search algorithm to find the value that minimized nonlinear GCV (Fu, 2005), which returned a value of 4.907.

### 3.4 Results

The coefficients for both the point process model and the MAXENT model (Web Table 1) are qualitatively different due largely to the different LASSO parameters. Of the 19 model coefficients, only 11 are nonzero in the point process model, as opposed to 17 for MAXENT. Moreover, the harsher LASSO penalty of the point process model ensures that each of the estimated coefficients are smaller than the corresponding coefficients of the MAXENT model. Otherwise, the models are broadly similar and hence the maps produced by both models identify the same geographic hot spots for *Corymbia eximia* (Figure 1b).

### 3.5 Summary

In analyzing the *Corymbia eximia* data we have seen a number of advantages of the Poisson point process model approach in choosing the spatial resolution, assessing model adequacy, and choosing the LASSO parameter. These are summarized in Table 1. Another potential advantage is in assessing model uncertainty—a point process framework can be used to put standard errors on model coefficients and predictions, although when using the LASSO in estimation (Fan and Li, 2001) there are some difficulties (Kyung et al., 2010). A final advantage worthy of mention is in computation time: Figure 1b took 12 seconds to produce for the point process model, but 130 seconds using MAXENT software (Table 1).

## 4. Improvements in Predictive Performance

We will now compare the predictive performance of the point process approach described in Section 3 to MAXENT in order to assess whether the refinements we proposed (in particular, modeling point interactions and data-driven estimation of the LASSO penalty parameter) improve the performance of the model. The approach we take is to model *Corymbia*

**Table 2**

Predictive performance (measured as average area under the ROC curve for 20 different fivefold spatial cross-validation schemes) of different presence-only models for *C. eximia* when predicting to a separate presence-absence dataset. Note that the point process approach proposed in Section 3 has the highest predictive performance

| Model            | LASSO penalty criteria | AUC    | Standard error |
|------------------|------------------------|--------|----------------|
| Poisson PPM      | No penalty             | 0.7555 | 0.0070         |
| MAXENT           | ad hoc MAXENT          | 0.8508 | 0.0060         |
| Poisson PPM      | Nonlinear GCV          | 0.8813 | 0.0051         |
| Area-interaction | Nonlinear GCV          | 0.9066 | 0.0036         |

*eximia* presence-only data and predict to new areas, assessing predictive performance using a separate presence-absence dataset from 8678 systematically collected transects (NSW Office of Environment and Heritage, 2010), as in Elith et al. (2006). This presence-absence dataset may be considered a “gold standard,” where observers have gone to each of the 8678 sites and specifically noted presences of *C. eximia*. We apply a spatial fivefold cross-validation in which sites are assigned to 30 square  $64 \times 64$  km spatial blocks that are randomly assigned to test and training samples. We employ this procedure to minimize the influence of spatial autocorrelation, which is not considered by MAXENT.

We evaluate the performance of MAXENT and various models from the point process approach by comparing predicted intensities at the systematically collected transects against observed presence/absence, using area under an ROC curve (Elith and Leathwick, 2007). Table 2 reveals that choosing the LASSO parameter to minimize the nonlinear GCV performs better than using MAXENT’s default method for *C. eximia* for both point process models. Hence, while MAXENT achieves high predictive performance relative to other methods (Elith et al., 2006), there is the potential to improve it further by using the data to inform the choice of the LASSO parameter.

## 5. Discussion

Some recent papers (Elith and Leathwick, 2009; Aarts, Fieberg, and Matthiopoulos, 2012) have called for greater unification and synthesis of the literature on SDM. To that end, we have demonstrated equivalence of MAXENT and a Poisson point process model. Warton and Shepherd (2010) showed the equivalence of Poisson point process models and pseudo-absence regression, which aside from MAXENT is the most commonly used approach to presence-only modeling at the moment. Hence our work represents a significant unification of the literature, using Poisson point process models to link the two most widely used presence-only methods, MAXENT and pseudo-absence regression. This work has significant practical ramifications, given that MAXENT (Table 1) and pseudo-absence regression (Warton and Shepherd, 2010) have shortcomings stemming largely from the framework used for modeling, which can be resolved by using a Poisson point process model instead. Others have made further connections between point process models and alternative approaches to

analysis—Aarts et al. (2012) made a connection to the estimation of “resource selection functions” via presence–absence analysis, and Dorazio (in press) to case-augmented binary regression. Point process models are a natural framework for analyzing presence-only data and it is interesting that a variety of different methods of analysis can all be connected to them in some way, and in many instances, improved through this connection.

A key distinction between point process models and MAXENT is that in the former we model  $\lambda(y)$  on a per area basis whereas for the latter, we model  $\pi(g_i)$  per grid cell—the per area analysis is thus invariant under choice of spatial resolution while the per grid cell analysis is not (because increasing spatial resolution increases the number of grid cells). This is related to the distinction between probability and frequency models (Aarts et al., 2012). It is this distinction that enables the likelihood convergence for a Poisson point process model (Figure 2a) and hence a data-driven choice of spatial resolution. However, MAXENT is proportional to a Poisson point process model (Theorem 2), which suggests that it can achieve the same qualitative answer but with the disadvantage of scale dependence of the predicted probabilities and an arbitrary choice of spatial resolution.

One important disadvantage of MAXENT is that in its current form, it does not estimate the intercept consistently (Elith et al., 2011). The intercept term diverges to  $-\infty$  as spatial resolution increases. Theorem 2 gives the form of the term causing this divergence. This means that MAXENT as currently posed cannot predict species intensity for any subset of the study region  $\mathcal{A}$  or likewise model abundance in the way that point process models can.

The new-found ability to use data to estimate spatial resolution (Figure 2a) is of interest for a couple of reasons. First, the resolution of the process is largely a function of biological factors and measurement error, and estimating this resolution informs us about the spatial scale at which such processes are operating. Second, the resolution of the process is of interest for computational reasons, because data are becoming available at increasingly fine resolutions—we originally had access to 8,620,092 points at the 100 m resolution, but even finer resolutions are now available—and analysis at such fine resolutions can be very computationally intensive. We know of colleagues analyzing this type of data in biology departments who have constructed their own parallel computing arrays to analyze this type of data for multiple species at fine resolutions. Hence it is of considerable practical interest to know whether such a fine resolution is required, and in our case, it clearly was not required as we only needed 134,716 quadrature points and were able to analyze data in seconds on a desktop computer (Table 1), with negligible loss of information.

An alternative approach to MAXENT analysis of all grid cells is to randomly select empty grid cells as “background points” for analysis. This obviates any computational need to coarsen resolution for analysis. The default approach that has been advocated (Phillips and Dudík, 2008) and implemented in MAXENT software is to use 10,000 random background points, which for our data was clearly insufficient (Figure 2a), equivalent to using a resolution of nearly 3 km. We advise that as a matter of routine, presence-only analysts should use their data to identify a spatial resolution appropriate for analysis,

or equivalently, to identify the number of “background points” to use in analysis.

In Section 4 we demonstrated that point process models achieve a higher predictive performance for *Corymbia eximia* by choosing the LASSO penalty parameter to minimize non-linear GCV. However, this may not be true of all species. We are currently investigating the question of how predictive performance varies with different methods of choosing the LASSO parameter across multiple species endemic to the Blue Mountains.

While MAXENT has become extraordinarily popular in ecology, the lack of a model-based framework and diagnostic tools means that it is used rather uncritically and sometimes inappropriately. A key advantage to the point process approach is that we have a model and can hence check assumptions. We found that a Poisson point process model did not fit *C. eximia* well due to violations of the independence assumption, in which case there are many alternative options for presence-only data (Baddeley and Turner, 2005; Chakraborty et al., 2011). The suite of diagnostic tools available via point process models (Cressie, 1993; Diggle, 2003; Baddeley and Turner, 2005; Baddeley et al., 2005) offers the possibility for users to think more critically about the appropriateness of the model they are fitting, which can ultimately have benefits in model interpretability and performance.

## 6. Supplementary Materials

Web Appendix 1, Web Figures 1 and 2 and Web Table 1 referenced in Sections 2 and 3 are available with this article at the Biometrics website on Wiley Online Library.

## ACKNOWLEDGEMENTS

We would like to thank Dan Ramp and Evan Webster for providing access to the data. We thank the Co-Editor, Associate Editor, and three referees for their valuable suggestions. IWR is supported by an Australian Postgraduate Award and DIW by the Australian Research Council Discovery Projects funding scheme (project number DP0985886).

## REFERENCES

- Aarts, G., Fieberg, J., and Matthiopoulos, J. (2012). Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods in Ecology and Evolution* **3**, 177–187.
- Baddeley, A. and Turner, R. (2005). Spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* **12**, 1–42.
- Baddeley, A. J. and van Lieshout, M. (1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* **47**, 601–619.
- Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* **54**, 329–350.
- Baddeley, A. J., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes. *Journal of the Royal Statistical Society, Series B* **67**, 617–666.
- Berman, M. and Turner, T. (1992). Approximating point process likelihoods with GLIM. *Journal of the Royal Statistics Society Series C – Applied Statistics* **41**, 31–38.

- Chakraborty, A., Gelfand, A., Wilson, A., Latimer, A., and Silander, J. (2011). Point pattern modelling for degraded presence-only data over large regions. *Journal of the Royal Statistical Society, Series C* **60**, 757–776.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. New York: John Wiley & Sons.
- Davis, P. J. and Rabinowitz, P. (1984). *Methods of Numerical Integration*, 2nd edition. Orlando, Florida: Academic Press, Inc.
- Della Pietra, S., Della Pietra, V., and Lafferty, J. (1997). Inducing features on random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**, 380–393.
- Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd edition. New York: Oxford University Press, Inc.
- Dorazio, R. M. (in press). Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*.
- Dutta, M. (1966). On maximum (information-theoretic) entropy estimation. *Sankhya: The Indian Journal of Statistics, Series A* **28**, 319–328.
- Elith, J. and Leathwick, J. R. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* **13**, 265–275.
- Elith, J. and Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**, 677–697.
- Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. McC., Peterson, A. T., Phillips, S. J., Richardson, K. S., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S., and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**, 129–151.
- Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* **77**, 802–813.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* **17**, 43–57.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**, 1348–1360.
- Fu, W. J. (2005). Nonlinear GCV and quasi-GCV for shrinkage models. *Journal of Statistical Planning and Inference* **131**, 333–347.
- Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables. *Annals of Mathematical Statistics* **34**, 911–934.
- Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Boca Raton, Florida: Chapman & Hall.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physics Review* **106**, 620–630.
- Kullback, S. (1959). *Information Theory and Statistics*. New York: John Wiley & Sons.
- Kyung, M., Gill, J., Ghosh, M., and Casella, G. (2010). Penalized regression, standard errors, and Bayesian lassos. *Bayesian Analysis* **5**, 369–411.
- McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, 2nd edition. London: Chapman and Hall.
- NSW Office of Environment and Heritage (2010). Atlas of NSW Wildlife database. Data accessed 20/04/2010.
- O'Sullivan, D. and Unwin, D. J. (2010). *Geographic Information Analysis*, 2nd edition. Hoboken, New Jersey: John Wiley & Sons.
- Pearce, J. L. and Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* **43**, 405–412.
- Phillips, S. J. and Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* **31**, 161–175.
- Phillips, S. J., Anderson, R. P., and Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**, 231–259.
- Thullier, W., Albert, C., Araújo, M., Berry, P., Cabeza, M., Guisan, A., Hickler, T., Midgley, G. F., Paterson, J., Schurr, F. M., Sykes, M. T., and Zimmerman, N. E. (2008). Predicting global change impacts on plant species distributions: Future challenges. *Perspectives in Plant Ecology, Evolution and Systematics* **9**, 137–152.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* **58**, 267–288.
- Warton, D. I. and 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.
- Zou, H., Hastie, T., and Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics* **35**, 2173–2192.

Received March 2012, Revised July 2012

Accepted August 2012.