

Hierarchical Species Distribution Models

Trevor J. Hefley^{1,2} · Mevin B. Hooten^{1,2,3}

Published online: 9 July 2016
© Springer International Publishing AG 2016

Abstract Determining the distribution pattern of a species is important to increase scientific knowledge, inform management decisions, and conserve biodiversity. To infer spatial and temporal patterns, species distribution models have been developed for use with many sampling designs and types of data. Recently, it has been shown that count, presence-absence, and presence-only data can be conceptualized as arising from a point process distribution. Therefore, it is important to understand properties of the point process distribution. We examine how the hierarchical species distribution modeling framework has been used to incorporate a wide array of regression and theory-based components while accounting for the data collection process and making use of auxiliary information. The hierarchical modeling framework allows us to demonstrate how several commonly used species distribution models can be derived from the point process distribution, highlight areas

of potential overlap between different models, and suggest areas where further research is needed.

Keywords Bayesian analysis · Citizen science · Count data · Presence-absence data · Presence-only data · Spatio-temporal statistics

Introduction

Mapping the world's biodiversity is a challenging responsibility. It is even more challenging to reliably predict how species' distributions will change due to threats such as climate change and habitat destruction [25, 26, 93]. To accomplish these important tasks, biogeographers, ecologists, statisticians, and the machine learning community have worked to create models that predict species' distributions and facilitate inference such as the habitat characteristics that individuals of a species prefer. The goal of this review is to present a unified view of many commonly used species distribution models. To accomplish this goal, we review the common types of data used to model species' distributions. We then discuss a unified spatio-temporal point process model for count, presence-absence, and presence-only data that has been developed within a hierarchical framework in several studies [11, 23, 31, 62]. The hierarchical modeling framework provides a way to incorporate important components of the data collection process (e.g., detection) and ecological processes (e.g., spread of an invasive species). By adding these components, the hierarchical framework allows for unification of many existing models and numerous extensions, including a wide variety of regression and theory-based models as well as explicit models of the data collection process.

This article is part of the Topical Collection on *Methodological Development*

✉ Trevor J. Hefley
Trevor.Hefley@colostate.edu

¹ Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO, USA

² Department of Statistics, Colorado State University, Fort Collins, CO, USA

³ Colorado Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Fort Collins, CO, USA

Data Types

There are three types of data commonly used for species distribution modeling: count, presence-absence, and presence-only. Although each type of data may seem distinct, we show in the next section that many types of data can be thought of as arising from a spatio-temporal point process (Fig. 1).

Count Data

Count data are generally collected using systematic surveys. Two systematic survey types that are used to generate count data are point counts and quadrat counts. Point counts involve at least one observer visiting a preselected point and recording the number of individuals they see during a prescribed time interval (e.g., Breeding Bird Survey data). Quadrat counts involve searching an area of known dimensions and recording the number of individuals found during a fixed period of time.

Presence-Absence Data

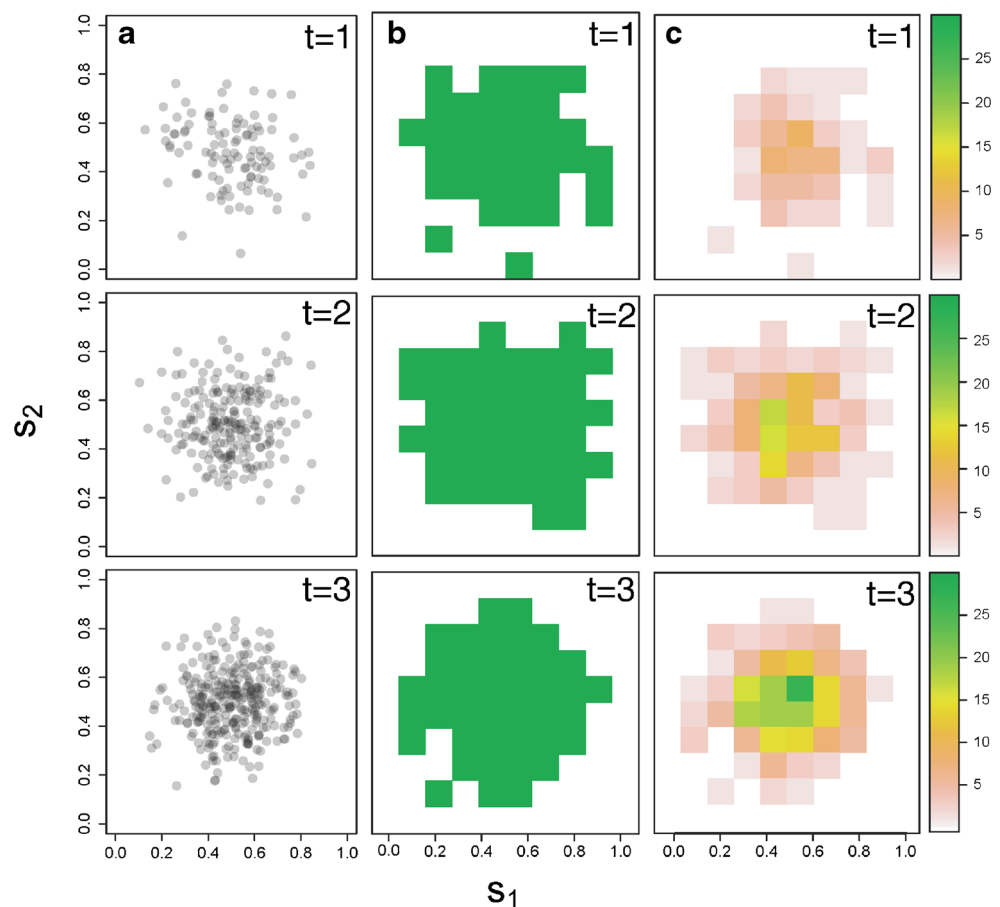
Presence-absence data can be count data that have been collapsed to binary responses where a one indicates that at least

one individual was present and a zero indicates absence or non-detection. Presence-absence data can also be collected using the same protocol as point counts and quadrat counts, but the survey is terminated when at least one individual has been observed.

Presence-Only Data

Unlike count and presence-absence data, which are usually collected at locations and time periods selected using a sampling design, presence-only data are typically collected opportunistically. Data sources and collation protocols can vary widely, but common sources include museum specimens and citizen contributed sightings of a species [7, 27, 46]. Typically presence-only data are reported as the location and time that an individual of a species was observed or collected. Another data collection process that gives rise to presence-only data is recording multiple locations of the same individual (e.g., telemetry data; [60]). For example, telemetry data is usually recorded as a time series of presence-only locations of an individual. The analysis of relocation data is often referred to as resource selection, but important connections between the statistical methods used for resource selection modeling and species distribution modeling exist [69, 73, 74].

Fig. 1 Discrete time inhomogeneous Poisson point pattern with an intensity ($\lambda(s, t)$) that increases at each time step (panel a; $t = 1, 2, 3$). Presence-absence data generated by aggregating the points into grid cells (panel b; $u = 0$ are white cells and $u = 1$ are green cells). Count data generated by aggregating points into grid cells and counting the total number of points in each cell (panel c)



Distribution Theory

An individual of a species exists as a point in space at any given time (Fig. 1a). A statistical model describing the pattern created by points in space and time is the space-time Poisson point process distribution [16, 20]. The log-likelihood for a space-time Poisson point process is

$$l(\lambda; \mathbf{U}) = \sum_{i=1}^n \log \lambda(\mathbf{s}_i, t_i) - \int_{\mathcal{A}} \int_0^T \lambda(\mathbf{s}, t) dt ds - \log(n!), \quad (1)$$

where the matrix \mathbf{U} is a $n \times 3$ matrix with rows that contain the location \mathbf{s}_i in two-dimensional space and time t_i of the i^{th} point ($i = 1, \dots, n$), \mathcal{A} is the two-dimensional study area, and the time period of observation is $[0, T]$. The intensity function $\lambda(\mathbf{s}, t)$ will depend on parameters and describes how the expected abundance (for an infinitely small region) changes within the study area and time period of observation. In words, Eq. 1 describes points that occur over a study area (\mathcal{A}) and time interval $[0, T]$ (Fig. 1a). The rate at which these points occur at any given location \mathbf{s} and point in time t is governed by the intensity function $\lambda(\mathbf{s}, t)$.

An important quantity derived from the point process model is the integrated intensity function

$$\bar{\lambda} = \int_{\mathcal{B}} \int_{t_1}^{t_2} \lambda(\mathbf{s}, t) dt ds, \quad (2)$$

where \mathcal{B} is an area of interest within the study area \mathcal{A} (i.e., $\mathcal{B} \subseteq \mathcal{A}$) over the time interval t_1 to t_2 ($0 \leq t_1 < t_2 \leq T$; hereafter \mathcal{B}, t_1 , and t_2 are termed the space-time volume of interest). An important result is that the number of points (u) within the space-time volume of interest follows a Poisson distribution

$$u \sim \text{Poisson}(\bar{\lambda}), \quad (3)$$

that depends on the integrated intensity function. Another important derived quantity is the probability that the space-time volume of interest has at least one point in it. This probability can be calculated as $P(u > 0) = 1 - e^{-\bar{\lambda}}$. It follows that the presence or absence of a species within the space-time volume of interest is distributed as

$$I(u > 0) \sim \text{Bernoulli}(1 - e^{-\bar{\lambda}}), \quad (4)$$

where the indicator function $I(u > 0)$ takes on a value of one if the volume has one or more points in it and zero otherwise.

The space-time Poisson point process distribution naturally describes presence-only data when recorded as the location and time an individual was observed or collected. As a result, variants of the Poisson point process model are popular for modeling species' distributions using presence-only data [83, 99]. By specifying an area (\mathcal{B}) and time

interval $[t_1, t_2]$ of interest, the space-time Poisson point process distribution can be linked to count and presence-absence data, which can be interpreted as points aggregated over a spatial and temporal extent ([1]; Fig. 1b & c). For example, \mathcal{B}, t_1 , and t_2 could represent a circle of a 100 m radius with an observer at the center counting the number of Bobwhite Quail (*Colinus virginianus*) seen within a 10 minute time interval. In this example, aggregation would occur when the observer records the count instead of the exact location and time that each Bobwhite was present. As another example, \mathcal{B}, t_1 , and t_2 could represent a 1 m² plot where the presence or absence of cheat grass (*Bromus tectorum*) is observed within a year. Regardless of whether the recorded data are counts, presence-absence, or presence-only, conceptualizing the data as a space-time Poisson point process is advantageous for the following reasons: 1) understanding the effects of spatial or temporal data aggregation (Fig. 1; 2) understanding and modeling the data collection process; and 3) preserving the ability to model continuous spatio-temporal dynamics.

Hierarchical Modeling Framework

Regardless of the type of data, in most studies, the quantity of interest is derived from the space-time intensity function $\lambda(\mathbf{s}, t)$. As we mentioned, $\lambda(\mathbf{s}, t)$ depends on parameters. For example, the intensity function is often formulated such that

$$\log(\lambda(\mathbf{s}, t)) = \mathbf{x}(\mathbf{s}, t)' \boldsymbol{\beta}, \quad (5)$$

where $\mathbf{x}(\mathbf{s}, t)$ is a $p \times 1$ vector that contains covariates at any given location and time within the study area and $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. Readers familiar with generalized linear models may recognize Eq. 5 as a linear predictor with a log link function (similar to that used with Poisson regression for count data). In fact, there are many important connections between certain generalized linear models (e.g., Poisson and logistic regression) and the Poisson point process model [1, 3, 32, 99]. Similar to the generalized linear modeling framework for count and binary data, Eqs. 1 and 5 specify a flexible model that can be used to describe how the response variable (e.g., a point pattern) changes over space and time due to covariates. The important connection is that count and binary data are aggregated point patterns that can be modeled using logistic or Poisson regression (Fig. 1), but interpreted as a discrete approximation to the space-time Poisson point process [1].

By placing the space-time Poisson point process model within a hierarchical modeling framework, the model can be made even more flexible. For example, a hierarchical modeling framework provides a way to incorporate important

components of both data collection and ecological processes. The framework we present originates from [8], but introductions can be found in [15, 16], and [51]. Following [35], we use the bracket notation to represent conditional probability density functions and write the hierarchical species distribution model as

$$g(\mathbf{Y}) \sim [g(\mathbf{Y})|\mathbf{U}, \boldsymbol{\theta}], \tag{6}$$

$$\mathbf{U} \sim [\mathbf{U}|\lambda(\mathbf{s}, t)], \tag{7}$$

$$\lambda(\mathbf{s}, t) \sim [\lambda(\mathbf{s}, t)|\boldsymbol{\beta}], \tag{8}$$

where Eq. 6 is called the data model and is a probability density function (or some transformation $g(\cdot)$ thereof) that describes the observed point level data \mathbf{Y} given the underlying spatio-temporal point pattern \mathbf{U} parameters $\boldsymbol{\theta}$. The observed point level data \mathbf{Y} may be aggregated to create binary or count data by transforming \mathbf{Y} with a deterministic function $g(\cdot)$. Equation 7 is called the sampling model, which is the space-time Poisson point process distribution from Eq. 1 that is conditional on the intensity

function $\lambda(\mathbf{s}, t)$. Finally, the distribution of $\lambda(\mathbf{s}, t)$ is called the process model and depends on the parameter vector $\boldsymbol{\beta}$ (Eq. 8). In Fig. 2, we show visual examples of Eqs. 6–8 with a directed acyclic graph (also called a Bayesian network) that can be used to understand and represent the hierarchical species distribution model [51]. In the following sections, we focus on each level of the hierarchical model and conclude with comments on implementation.

Process Model

The process model (Eq. 8) can take a wide range of forms depending on the goals of the study. For example, the process model may have a relatively simple form such as a log-linear model that includes covariates $\mathbf{x}(\mathbf{s}, t)$ (e.g., Eq. 5) or it may be a dynamic model such as a partial differential equation. There are three common classes of process models used to determine the distribution of a species: parametric regression, semiparametric regression, and dynamic spatio-temporal models.

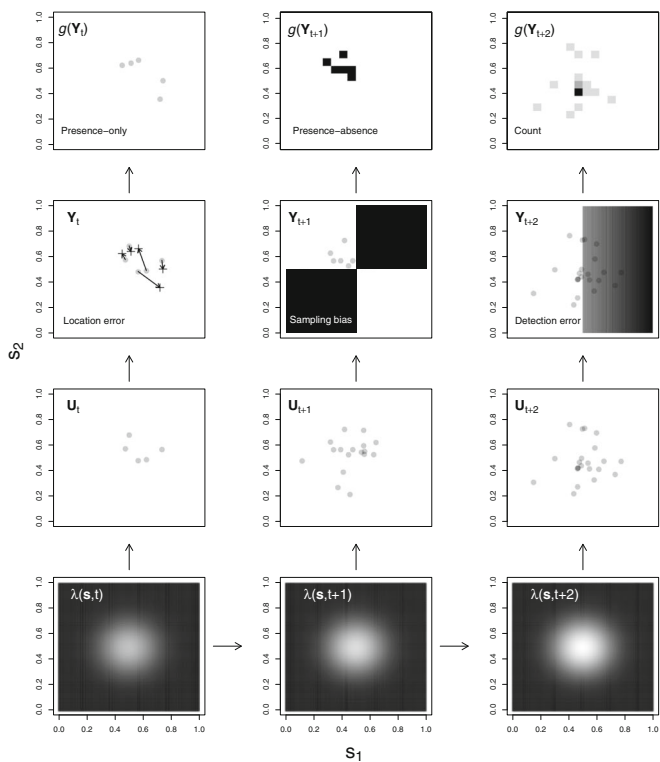
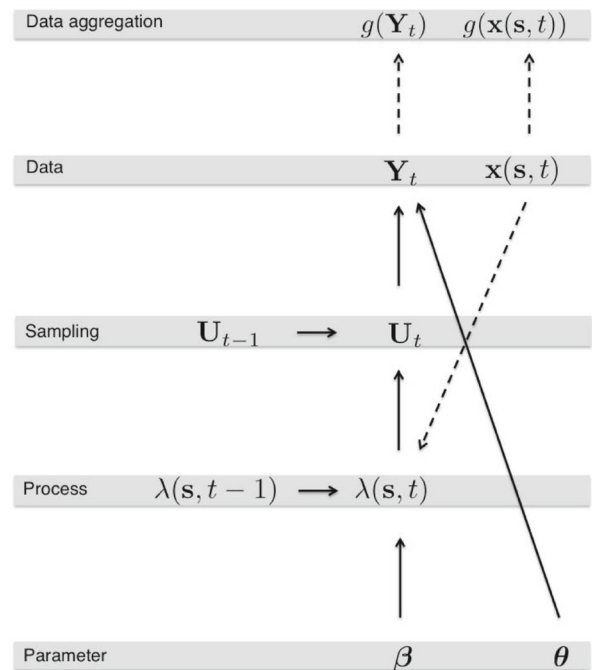


Fig. 2 Example aggregated data, observed point data, true point data, and processes (left figures) that can be used with the directed acyclic graph (right; arrow diagram) to visualize and represent a hierarchical species distribution model. Point pattern data \mathbf{Y} at three time steps that are observed with error when compared to the underlying true point pattern \mathbf{U} . The point pattern data \mathbf{Y} might be aggregated by a transformation $g(\cdot)$ during the data collection process (e.g., points represented as count or presence-absence in grid cells). The true point pattern (\mathbf{U}) depends on the spatio-temporal intensity function $\lambda(\mathbf{s}, t)$ which was



generated using a spatio-temporal diffusion process. The data \mathbf{Y}_t were observed with location error where the gray points are the true locations and black + signs are the recorded locations. The data \mathbf{Y}_{t+1} were observed from biased sampling efforts that resulted in the blacked out areas not being sampled. The data \mathbf{Y}_{t+2} resulted from non-detection of \mathbf{U}_{t+2} which is represented by the difficulty of observing points in the shaded region. Note, in the directed acyclic graph, the solid lines show stochastic relationships and the dashed lines show deterministic relationships

Parametric Regression

The goal of parametric regression is to formulate a link between abundance of a species and covariates. This is typically accomplished using a model of the form $\log(\lambda(\mathbf{s}, t)) = \mathbf{x}(\mathbf{s}, t)' \boldsymbol{\beta}$, where the log function links the known spatial covariates and unknown regression parameters $\boldsymbol{\beta}$ to the underlying intensity $\lambda(\mathbf{s}, t)$. Similar to generalized linear models, the strength of the parametric regression model is that it is simple to implement and easy to interpret.

Semiparametric Regression

As a generalization of parametric regression, semiparametric regression also links the underlying intensity $\lambda(\mathbf{s}, t)$ to some number of covariates $\mathbf{x}(\mathbf{s}, t)$. The difference is that semiparametric regression involves functions of the covariates that allow the model to capture nonlinear effects (e.g., polynomial regression). In most situations, semiparametric regression is implemented using basis functions [47]. Many of the statistical and machine learning methods used for species distribution modeling rely on basis functions. For example, generalized additive models are used to model species' distributions and employ various basis functions to model the smooth effect of a covariate [39, 44]. Other commonly used methods that rely on basis functions include regression trees [29, 44], some maximum entropy methods (e.g., MAXENT; [30]), and latent Gaussian process models [12, 21, 37, 53, 59, 67]. One advantage of semiparametric regression is that smooth effects of the spatial location or time period can be used to account for spatial or temporal autocorrelation [47].

Dynamic Process Models

Many ecological theories have generated hypotheses about why the abundance of a species varies across space and through time [52]. However, the linkages between species distribution modeling approaches based on regression or semiparametric regression models and ecological theory are weak [24, 28, 40]. Dynamic spatio-temporal models can be used within the hierarchical framework to link theoretical models to data when modeling the distribution of a species. For example, species distribution models are commonly used to understand and predict the spread of an invasive species. A simple partial differential equation that describes the spread (ecological diffusion) and logistic growth of a population is

$$\frac{\partial \lambda(\mathbf{s}, t)}{\partial t} = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \right) \delta(\mathbf{s}, t) + r \left(1 - \frac{\lambda(\mathbf{s}, t)}{k} \right), \quad (9)$$

where s_1 and s_2 are elements of the spatial coordinate \mathbf{s} and $\delta(\mathbf{s}, t)$ is the diffusion coefficient (or motility coefficient) that controls the rate of spread and is inversely related to residence time [34]. The parameters r and k are analogous to the growth rate and equilibrium population size in the traditional logistic growth model when expressed as an ordinary differential equation. Various implementations of this approach have been used to model the invasion and spread of Eurasian Collared-Doves (*Streptopelia decaocto*), House Finches (*Carpodacus mexicanus*), Common Mynas (*Acridotheres tristis*), and mountain pine beetles (*Dendroctonus ponderosae*) [9, 55, 82, 101, 104]. Partial differential equation models are well developed for many ecological processes and are easy to modify to match the goals of a study [52]; any of the parameters in Eq. 9 could depend on covariates $\mathbf{x}(\mathbf{s}, t)$ (e.g., $\delta(\mathbf{s}, t) = \mathbf{x}(\mathbf{s}, t)' \boldsymbol{\beta}$). For example, [54] allowed the diffusion coefficient $\delta(\mathbf{s}, t)$ to depend on human population density and showed that diffusion of Eurasian Collared-Doves might be negatively associated with human population density.

Dynamic statistical models enable scientific inference that can not be obtained when using parametric or semiparametric regression [102]. The dynamical approach to statistical modeling has been present in the ecological literature for quite some time [101] and has been used in many areas of environmental research [16], but has not been widely used for species distribution modeling.

Sampling Models

The sampling model (Eq. 7) is a probability density function that describes the data that would be observed if the data could be recorded perfectly. For the space-time Poisson point process model, the matrix \mathbf{U} (where each row is the exact location and time that an individual was present) contains the data that would have been recorded if there were no errors in observation or aggregation of the data (Fig. 2). For most contemporary analyses of presence-only data, the Poisson point process is the sampling model that has been used and advocated as the preferred method by many authors [1, 43, 83, 84, 98, 99]. Typically, count data and presence-absence data are treated differently due to the aggregation that occurs during the data collection process (Fig. 1; [1]). Given the space-time Poisson point process, the aggregate (over space and time) of \mathbf{U} is \mathbf{u} , a $m \times 1$ vector that contains the true count or presence-absence for m discrete sample units during some time interval (Fig. 1b & c). Associated with each element of \mathbf{u} is the location of the discrete sampling unit and the time period it was sampled. As shown in the distribution theory section, aggregation of a spatio-temporal Poisson point process results in Poisson or Bernoulli distribution for \mathbf{u} . The interpretation when using

Poisson or Bernoulli sampling models can be linked to the Poisson point process via the integrated intensity function (Eq. 2).

Data Models

In almost all studies, the underlying spatio-temporal point pattern \mathbf{U} (or some aggregate \mathbf{u}) is not observed perfectly (e.g., [5, 61, 65, 66, 77]; Fig. 2). The hierarchical modeling framework provides a natural way to incorporate uncertainty in observations. This is accomplished via the data model (Eq. 6) which is a conditional distribution that describes the observed data \mathbf{Y} (or some aggregate $g(\mathbf{Y}) \equiv \mathbf{y}$) given the true underlying point pattern \mathbf{U} (Fig. 2). For most species distribution modeling efforts, determining which data model is appropriate can be the greatest challenge to obtaining reliable predictions and inference. Next, we highlight common data models for count, presence-absence, and presence-only data.

Data Models for Counts

Count data generally suffer from two types of observer error. Within an area of interest \mathcal{B} , over the time period t_1 to t_2 , the true number of individuals present is u . The observed count y can contain individuals that were counted multiple times, not detected, or both. The most common data model assumes that the only observation error is non-detection. With a Poisson sampling model, this approach is called the N-mixture model [85]

$$y|u, p \sim \begin{cases} \text{Binomial}(u, p) & , u > 0 \\ 0 & , u = 0 \end{cases} \quad (10)$$

where y is the observed count, u is the unknown true number of individuals, and p is the probability of detection. There are many variations of the N-mixture model, but most implementations require repeated sampling over multiple time periods during which the true number of individuals present (u) within the area (\mathcal{B}) and time period ($[t_1, t_2]$) of interest can be assumed to be constant. The implication of this assumption is that the underlying point pattern must be static over the multiple sampling intervals to ensure that u is constant. There are alternative models that require only a single sample [17, 50, 90], but such models require stronger assumptions or structure in the process model and are a current topic of debate [63, 64, 91].

Data Models for Presence-Absence

Presence-absence data generally suffer from similar observation errors as count data, but these errors are termed

false-positive and false-negatives (non-detection). Data models for false negatives are the most widely used (e.g., [71, 95]). A commonly used form of the data model for false-negatives is

$$y|u, p \sim \begin{cases} \text{Bernoulli}(p) & , u > 0 \\ 0 & , u = 0 \end{cases} \quad (11)$$

where y is the observed presence or absence, u is the unknown true number of individuals, and p is the probability that at least one individual is detected. As with the N-mixture model, Eq. 11 requires repeated sampling over multiple time periods during which the occupancy is constant (i.e., $u > 0$ or $u = 0$ must remain constant). Numerous variations of Eq. 11 have been used to account for false-negatives that make less restrictive assumptions (e.g., data from single-site visits [18, 70]). Recently there have been controlled experiments to assess the performance of these data models when basic assumptions are not met (e.g., the assumption that no false positives occur [75]). Also, data models have been recently developed to deal with false-positives [2, 13, 87, 88] and species misidentification [76]. Although Eq. 11 and variants have been widely used as data models, there is debate about whether the model assumptions are met in practice [38, 100].

Data Models for Presence-Only

Unlike count and presence-absence data, presence-only data are almost always collected opportunistically. Determining what types of observational errors might have occurred with presence-only data and how these errors can be modeled is an emerging area of research. Three common types of observational errors that we discuss are: sampling bias, non-detection, and location error (Fig. 2).

As before, let \mathcal{A} be the two-dimensional “study area” and \mathcal{B} be an area that was “sampled.” With count and presence-absence data, we know \mathcal{B} as a result of the survey design. For most designed surveys, there will be multiple areas that are sampled so that we have $\mathcal{B}_1, \dots, \mathcal{B}_m$ non-overlapping sites. As a result, the entire study area (\mathcal{A}) is the union of the areas that were sampled ($\mathcal{A} = \mathcal{B}_1 \cup \dots \cup \mathcal{B}_m$; note that a similar argument applies to the time interval $[0, T]$). For designed surveys, \mathcal{A} is a known spatial domain, but for presence-only data there is rarely information to determine \mathcal{A} exactly. The space-time Poisson point process model requires that \mathcal{A} is known due to the limits of integration in Eq. 1. Various *ad hoc* methods have been used in an attempt to remedy the fundamental lack of information that would otherwise have been obtained from the survey design. For example, a common technique is to sample a large number (q) of randomly (or uniformly) located points where it is assumed that the species was not present

(often called pseudo-absences). The area where the pseudo-absences are sampled from (\mathcal{P}) generally encompasses the locations of the observed data. The presence-only data and pseudo-absences are then treated as presence-absence data. Alternatively, Warton and Shepherd [99] showed that the pseudo-absences are a numerical technique for approximating the integral in Eq. 1, if one assumes that $\mathcal{A} = \mathcal{P}$. Numerous studies have explored the bias from the so-called pseudo-absence approach for binary data. Given the connection between binary data and the Poisson point process models (Eq. 4), the results from studies on the pseudo-absence approach for binary data should also apply to the Poisson point process model (e.g., [4, 80, 103] and many others). How the study area \mathcal{A} is defined has the potential to influence both prediction and inference [6, 96]. However, for most situations the assumption that $\mathcal{A} = \mathcal{P}$ is difficult to assess. We are not aware of a data model that can remedy the problems associated with this form of sampling bias.

Similar to count and presence-absence data, some individuals go undetected. As before, let \mathcal{A} be the two-dimensional study area. Now assume that everywhere within \mathcal{A} there is a non-zero probability that a point could be detected ($p(\mathbf{s}, t)$). In effect, we are assuming that the entire study area was “sampled,” and that all individuals could have been detected. This assumption differs from that in the previous example in that now we are assuming that \mathcal{A} is known, but that not all points within \mathcal{A} are detected. For example, [46] presented data that included citizen contributed sightings of the endangered Whooping Crane (*Grus americana*) within the state of Nebraska, USA. For the study presented in [46], it might be reasonable to assume that \mathcal{A} is the state of Nebraska and that detection was not perfect. Assume, for a moment, that $p(\mathbf{s}, t) = 1, \forall \mathbf{s}, t$. As a result, detection of points would be perfect and the observed data would be \mathbf{U} (the $n \times 3$ matrix with rows that contains the location and time of each sighting; Fig. 2). If detection is not perfect, then the observed data are \mathbf{Y} (the $m \times 3$ matrix with rows that contains the location and time of each observed individual where $m \leq n$). Non-detection results in a thinned spatio-temporal Poisson point process model, which has the following likelihood

$$l(\lambda, p; \mathbf{Y}) = \sum_{i=1}^m \log \lambda(\mathbf{s}_i, t_i) p(\mathbf{s}, t) - \int_{\mathcal{A}} \int_0^T \lambda(\mathbf{s}, t) p(\mathbf{s}, t) dt ds - \log(m!). \tag{12}$$

The thinned Poisson point process (Eq. 12) is the same as Eq. 1, except that the intensity function $\lambda(\mathbf{s}, t)$ has been multiplied by (or is a convolution with) the probability of detection $p(\mathbf{s}, t)$. To understand the implications of thinning a point process, let \mathcal{B} be an area of interest within the

study area \mathcal{A} over the time interval $[t_1, t_2]$ and define the integrated probability of detection as

$$\bar{p} = \frac{1}{|\mathcal{B}| |t_2 - t_1|} \int_{\mathcal{B}} \int_{t_1}^{t_2} p(\mathbf{s}, t) dt ds, \tag{13}$$

where $|\mathcal{B}|$ is the area of \mathcal{B} and $|t_2 - t_1|$ is length of the time interval of interest. As a result, the true number of individuals within \mathcal{B} over the time interval $[t_1, t_2]$ is u as in Eq. 3, but the observed number of individuals is distributed

$$y \sim \text{Poisson}(\bar{p} \bar{\lambda}), \tag{14}$$

where $\bar{\lambda}$ is the integrated intensity from Eq. 2. When only \mathbf{Y} (or \mathbf{y}) is observed, and without additional information about $\lambda(\mathbf{s}, t)$ or $p(\mathbf{s}, t)$ (or any derived quantities $\bar{\lambda}$ and \bar{p}), it is impossible to simultaneously estimate both quantities. Without a correction for non-detection, parameters estimated from the process model for $\lambda(\mathbf{s}, t)$ will be biased. There are special cases when using parametric regression process models where the bias might not be an issue [22, 49, 98], but it is unlikely that the required assumptions will be met in practice and are challenging to verify in most cases [49]. To correct for bias, recent efforts have paired presence-only data with additional data sources such as expert knowledge [46] or a smaller number of perfectly detected presence-absence or count data [23, 31, 33, 36, 42]. Determining what sources of data can be used to correct for the bias and how to incorporate auxiliary sources of data is an ongoing area of research.

Another common source of observational error occurs when the individual is present at a location that is different than the recorded location (Fig. 2). For some historical records, the recorded location may be the nearest road, town, or other feature on the landscape. For many applications, this presents a problem because $\lambda(\mathbf{s}, t)$ varies spatially. As a result, the intensity at the recorded locations will be different than at the true locations. For example, location error is a problem when $\lambda(\mathbf{s}, t)$ depends on spatial covariates (e.g., Eq. 5) and the covariates at the recorded location are different than the covariates at the true location [45]. With respect to count and presence-absence data arising from a space-time Poisson point process, an implicit assumption is that the covariates can be represented by a single point within the discrete sampling units. Aggregation introduces location error, and it is important to check whether this error has a noticeable effect on inference [31]. Development of data models that account for location error within the hierarchical modeling framework is needed. Data models developed for accommodating covariate measurement error may provide useful extensions to correct for location error [45, 92].

Implementation

Hierarchical models can be implemented under a Bayesian or likelihood paradigm (e.g., [51, 68, 81, 86]). The complexity of the model usually dictates which paradigm is used. For example, it is common to implement hierarchical models that use partial differential equation (e.g., Eq. 9) using Bayesian algorithms such as Markov chain Monte Carlo (MCMC; [102]) or integrated nested Laplace approximation [57, 58], whereas, simpler models that use basis functions might use penalized maximum likelihood [47].

Data models are rarely used in the analysis of presence-only data (but see [23, 31, 45]). If sampling error is ignored, the Poisson point process model is relatively straightforward to implement using maximum likelihood, penalized maximum likelihood, or Bayesian estimation (via MCMC or integrated nested Laplace approximation [57, 58]). Recently there has been a focus on understanding how generalized linear models with Poisson or Bernoulli response distributions can be used to approximate the Poisson point process model [1, 3, 32, 99]. Most software used to implement semiparametric regression was developed as an extension of the generalized linear modeling framework (e.g., generalized additive models). A major benefit of the recent work on approximating the Poisson point process model is that widely available software can be used when the data are assumed to be observed without error [32].

Regardless of the method used, an important task when implementing the point process model is to approximate the integral in Eq. 1 [99], which requires spatial covariate data at all locations within the study area. The quality of this approximation can be assessed by viewing plots of the estimated likelihood or parameters against the number of integration points (e.g., see Fig. 1 in [78], Fig. 2 in [83], or Appendix S2 in [48]). Also important to implementation are various computational problems, such as infinite and boundary estimates, that can also arise for particular species distribution models and data sets [19, 41, 48, 100]. At present, we are unaware of a comprehensive review of the various estimation methods for implementing hierarchical species distribution models. Thus, an awareness of the variety of options available to implement hierarchical models is helpful so that the chosen method will best match the goals of the study.

Conclusion

Hierarchical models are widely used in ecology [14, 15]. In practice, the hierarchical modeling framework has been particularly useful when implementing process models that incorporate ecological theory or when modeling the data collection process (e.g., [54, 55, 101]). The hierarchical

modeling framework has also been useful and widely implemented for count and presence-absence data [72, 86], but has not been readily used for presence-only data (but see [23, 31]). Although this review is limited to hierarchical models for a single species, the hierarchical species distribution modeling approach can be readily extended to include multiple, and possibly interacting, species [56, 79, 97]. Given the success for analyzing other types of data, we believe the hierarchical modeling framework has great promise for presence-only data.

Although many data models exist for count and presence-absence data, there are relatively few explicit data models that have been developed for presence-only data. In many cases, the data collection process is ignored, which might be due to the complex and often unknown observation process associated with opportunistic presence-only data. Opportunistic data collation may result in very complicated errors that are difficult to model without auxiliary data or strong assumptions [22, 23, 31, 45, 46, 49, 94, 98]. The key to obtaining reliable inference from opportunistic presence-only data is determining which additional data sources are needed and available to responsibly model the data collection process. Recently, the spatio-temporal Poisson point process model has been used for the analysis of animal movement data collected using telemetry devices [10, 60, 89]. We anticipate that data models developed for telemetry data will have an analogous use for species distribution models (e.g., [10]). We also anticipate that the methodology developed to account for repeated measurements (i.e., locations) of the same individual(s) developed for telemetry data, will have analogous use for species distribution models that are used to model count, presence-absence, and presence-only data that includes multiple observations of the same individual(s).

Acknowledgments We thank Robert Dorazio, John Fieberg, and Ashley Gramza for valuable insight and discussions about this work. The authors acknowledge support for this research from NSF 1614392 and USGS G14AC00366. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

1. Aarts G, Fieberg J, Matthiopoulos J. Comparative interpretation of count, presence-absence and point methods for species distribution models. *Methods Ecol Evol.* 2012;3(1):177–87.

2. Aing C, Halls S, Oken K, Dobrow R, Fieberg J. A Bayesian hierarchical occupancy model for track surveys conducted in a series of linear, spatially correlated, sites. *J Appl Ecol*. 2011;48(6):1508–17.
3. Baddeley A, Berman M, Fisher N, Hardegen A, Milne R, Schuhmacher D, et al. Spatial logistic regression and change-of-support in Poisson point processes. *Electronic Journal of Statistics*. 2010;4:1151–1201.
4. Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. Selecting pseudo-absences for species distribution models: how, where and how many. *Methods Ecol Evol*. 2012;3(2):327–38.
5. Barry S, Elith J. Error and uncertainty in habitat models. *J Appl Ecol*. 2006;43(3):413–23.
6. Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, et al. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecol Model*. 2011;222(11):1810–19.
7. Belaire J, Kreakie BJ, Keitt T, Minor E. Predicting and mapping potential whooping crane stopover habitat to guide site selection for wind energy projects. *Conserv Biol*. 2014;28(2):541–50.
8. Berliner LM. Hierarchical Bayesian time series models. In: *Maximum Entropy and Bayesian Methods*. Springer; 1996. p. 15–22.
9. Broms KM, Hooten MB, Johnson DS, Altwegg R, Conquest LL. Dynamic occupancy models for explicit colonization processes. *Ecology*. 2016;97(1):194–204.
10. Brost BM, Hooten MB, Hanks EM, Small RJ. Animal movement constraints improve resource selection inference in the presence of telemetry error. *Ecology*. 2015;96:2590–97.
11. Chakraborty A, Gelfand AE, Wilson AM, Latimer AM, Silander JA. Point pattern modelling for degraded presence-only data over large regions. *J R Stat Soc: Ser C: Appl Stat*. 2011;60(5):757–76.
12. Chakraborty A, Gelfand AE, Wilson AM, Latimer AM, Silander Jr JA. Modeling large scale species abundance with latent spatial processes. *Ann Appl Stat*. 2010;28(4):1403–29.
13. Chambert T, Miller DA, Nichols JD. Modeling false positive detections in species occurrence data under different study designs. 2015;96(1):332–39.
14. Clark JS. Why environmental scientists are becoming Bayesians. *Ecol Lett*. 2005;8(1):2–14.
15. Cressie N, Calder CA, Clark JS, Ver Hoef JM, Wikle CK. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol Appl*. 2009;19(3):553–70.
16. Cressie N, Wikle C. *Statistics for Spatio-Temporal Data*: Wiley; 2011.
17. Dail D, Madsen L. Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*. 2011;67(2):577–87.
18. Dail D, Madsen L. Estimating open population site occupancy from presence–absence data lacking the robust design. *Biometrics*. 2013;69(1):146–56.
19. Dennis EB, Morgan BJ, Ridout MS. Computational aspects of N-mixture models. *Biometrics*. 2015;71(1):237–46.
20. Diggle PJ. *Statistical analysis of spatial and spatio-temporal point patterns*: CRC Press; 2013.
21. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Stat Sci*. 2013;4(3):542–63.
22. Dorazio RM. Predicting the geographic distribution of a species from presence-only data subject to detection errors. *Biometrics*. 2012;68(4):1303–12.
23. Dorazio RM. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Glob Ecol Biogeogr*. 2014;23(12):1472–84.
24. Dormann CF, Schymanski SJ, Cabral J, Chuine I, Graham C, Hartig F, et al. Correlation and process in species distribution models: bridging a dichotomy. *J Biogeogr*. 2012;39(12):2119–31.
25. Duputié A, Zimmermann NE, Chuine I. Where are the wild things? Why we need better data on species distribution. *Glob Ecol Biogeogr*. 2014;23(4):457–67.
26. Elith J, Kearney M, Phillips S. The art of modelling range-shifting species. *Methods Ecol Evol*. 2010;1(4):330–42.
27. Elith J, Leathwick J. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Divers Distrib*. 2007;13(3):265–75.
28. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst*. 2009;40(1):677.
29. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol*. 2008;77(4):802–13.
30. Elith J, Phillips SJ, Hastie T, Dudík M, Chee YE, Yates CJ. A statistical explanation of MAXENT for ecologists. *Divers Distrib*. 2011;17(1):43–57.
31. Fithian W, Elith J, Hastie T, Keith DA. Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods Ecol Evol*. 2015;6(4):424–38.
32. Fithian W, Hastie T. Finite-sample equivalence in statistical models for presence-only data. *Ann Appl Stat*. 2013;7(4):1917–39.
33. Fletcher RJ, McCleery RA, Greene DA, Tye C. Integrated models that unite local and regional data reveal larger-scale environmental relationships and improve predictions of species distributions. *Landsc Ecol*.
34. Garlick MJ, Powell JA, Hooten MB, McFarlane LR. Homogenization of large-scale movement models in ecology. *Bull Math Biol*. 2011;73(9):2088–2108.
35. Gelfand AE, Smith AF. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc*. 1990;85(410):398–409.
36. Giraud C, Calenge C, Coron C, Julliard R. Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics*. 2016;72(2):649–58.
37. Golding N, Purse BV. Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Meth Ecol Evol*. 2016;7(5):598–608.
38. Guillera-Arroita G, Lahoz-Monfort JJ, MacKenzie DI, Wintle BA, McCarthy MA. Ignoring imperfect detection in biological surveys is dangerous: A response to 'fitting and interpreting occupancy models'. *PLoS One*. 2014;9(7):e99571.
39. Guisan A, Edwards TC, Hastie T. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol Model*. 2002;157(2):89–100.
40. Guisan A, Thuiller W. Predicting species distribution: offering more than simple habitat models. *Ecol Lett*. 2005;8(9):993–1009.
41. Haines LM. (in press). Maximum likelihood estimation for N-mixture models. *Biometrics*.
42. Hanks EM, Hooten MB, Baker FA. Reconciling multiple data sources to improve accuracy of large-scale prediction of forest disease incidence. *Ecol Appl*. 2011;21(4):1173–88.
43. Hastie T, Fithian W. Inference from presence-only data; the ongoing controversy. *Ecography*. 2013;36(8):864–67.
44. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition: Springer Series in Statistics. Springer; 2009.

45. Hefley TJ, Baasch DM, Tyre AJ, Blankenship EE. Correction of location errors for presence-only species distribution models. *Methods Ecol Evol.* 2014;5(3):207–14.
46. Hefley TJ, Baasch DM, Tyre AJ, Blankenship EE. Use of opportunistic sightings and expert knowledge to predict and compare whooping crane stopover habitat. *Conserv Biol.* 2015;29:1337–46.
47. Hefley TJ, Broms KM, Brost BM, Buderman FE, Kay SL, Scharf HR, et al. The basis function approach for modeling autocorrelation in ecological data. [arXiv:1606.05658](https://arxiv.org/abs/1606.05658).
48. Hefley TJ, Hooten MB. On the existence of maximum likelihood estimates for presence-only data. *Methods Ecol Evol.* 2015;6:648–55.
49. Hefley TJ, Tyre AJ, Baasch DM, Blankenship EE. Nondetection sampling bias in marked presence-only data. *Ethol Evol.* 2013;3(16):5225–36.
50. Hefley TJ, Tyre AJ, Blankenship EE. Fitting population growth models in the presence of measurement and detection error. *Ecol Model.* 2013;263:244–50.
51. Hobbs NT, Hooten MB. *Bayesian Models A Statistical Primer for Ecologists.* Princeton University Press. 2015.
52. Holmes EE, Lewis MA, Banks J, Veit R. Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology.* 1994;157:17–29.
53. Hooten MB, Larsen DR, Wikle CK. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landsc Ecol.* 2003;18(5):487–502.
54. Hooten MB, Wikle CK. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian collared-dove. *Environ Ecol Stat.* 2008;15(1):59–70.
55. Hooten MB, Wikle CK, Dorazio RM, Royle JA. Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics.* 2007;63(2):558–67.
56. Hui FK. *boral*—Bayesian ordination and regression analysis of multivariate abundance data in R. *Methods in Ecology and Evolution.* 2016;7(5):744–50.
57. Illian JB, Martino S, Sørbye SH, Gallego-Fernández JB, Zunzunegui M, Esquivias MP, et al. Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods Ecol Evol.* 2013;4(4):305–15.
58. Illian JB, Soerbye S, Rue H, Hendrichsen D. Using INLA to fit a complex point process model with temporally varying effects—a case study. *Journal of Environmental Statistics.* 2012.
59. Johnson DS, Conn PB, Hooten MB, Ray JC, Pond BA. Spatial occupancy models for large data sets. *Ecology.* 2013;94(4):801–08.
60. Johnson DS, Hooten MB, Kuhn CE. Estimating animal resource selection from telemetry data using point process models. *J Anim Ecol.* 2013;82(6):1155–64.
61. Kéry M. Towards the modelling of true species distributions. *J Biogeogr.* 2011;38(4):617–18.
62. Kéry M, Royle JA. *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS.* Academic Press. 2016.
63. Knappe J, Korner-Nievergelt F. Estimates from non-replicated population surveys rely on critical assumptions. *Methods Ecol Evol.* 2015;6(3):298–306.
64. Knappe J, Korner-Nievergelt F. On assumptions behind estimates of abundance from counts at multiple sites. *Methods Ecol Evol.* 2016;7(2):206–09.
65. Kramer-Schadt S, Niedballa J, Pilgrim JD, Schröder B, Lindenborn J, Reinfelder V, et al. The importance of correcting for sampling bias in MAXENT species distribution models. *Divers Distrib.* 2013;19(11):1366–79.
66. Lahoz-Monfort JJ, Guillera-Aroita G, Wintle BA. Imperfect detection impacts the performance of species distribution models. *Glob Ecol Biogeogr.* 2014;23(4):504–15.
67. Latimer AM, Wu S, Gelfand AE, Silander Jr JA. Building statistical models to analyze species distributions. *Ecol Appl.* 2006;16(1):33–50.
68. Lele SR, Dennis B, Lutscher F. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecol Lett.* 2007;10(7):551–63.
69. Lele SR, Merrill EH, Keim J, Boyce MS. Selection, use, choice and occupancy: clarifying concepts in resource selection studies. *J Anim Ecol.* 2013;82(6):1183–91.
70. Lele SR, Moreno M, Bayne E. Dealing with detection error in site occupancy surveys: what can we do with a single survey?. *J Plant Ecol.* 2012;5(1):22–31.
71. MacKenzie DI, Nichols JD, Lachman GB, Droege S, Andrew Royle J, Langtimm CA. Estimating site occupancy rates when detection probabilities are less than one. *Ecology.* 2002;83(8):2248–55.
72. MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LL, Hines JE. *Occupancy Estimation and Modeling Inferring Patterns and Dynamics of Species Occurrence.* Academic Press. 2006.
73. McDonald L, Manly B, Huettmann F, Thogmartin W. Location-only and use-availability data: analysis methods converge. *J Anim Ecol.* 2013;82(6):1120–24.
74. McDonald TL. The point process use-availability or presence-only likelihood and comments on analysis. *J Anim Ecol.* 2013;82(6):1174–82.
75. Miller DA, Bailey LL, Grant Campbell EH, McClintock BT, Weir LA, Simons TR. Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods Ecol Evol.* 2015;6(5):557–65.
76. Miller DA, Nichols JD, McClintock BT, Grant Campbell EH, Bailey LL, Weir LA. Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology.* 2011;92(7):1422–28.
77. Monk J. How long should we ignore imperfect detection of species in the marine environment when modelling their distribution?. *Fish Fish.* 2014;15(2):352–58.
78. Northrup JM, Hooten MB, Anderson CR, Wittemyer G. Practical guidance on characterizing availability in resource selection functions under a use-availability design. *Ecology.* 2013;94(7):1456–63.
79. Ovaskainen O, Soininen J. Making more out of sparse data: hierarchical modeling of species communities. *Ecology.* 2011;92(2):289–95.
80. Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol Appl.* 2009;19(1):181–97.
81. Ponciano JM, Taper ML, Dennis B, Lele SR. Hierarchical models in ecology: confidence intervals, hypothesis testing, and model selection using data cloning. *Ecology.* 2009;90(2):356–62.
82. Powell JA, Bentz BJ. Phenology and density-dependent dispersal predict patterns of mountain pine beetle (*dendroctonus ponderosae*) impact. *Ecol Model.* 2014;273:173–85.
83. Renner IW, Elith J, Baddeley A, Fithian W, Hastie T, Phillips SJ, et al. Point process models for presence-only analysis. *Methods Ecol Evol.* 2015;6(4):366–79.
84. Renner IW, Warton DI. Equivalence of MAXENT and Poisson point process models for species distribution modeling in ecology. *Biometrics.* 2013;69(1):274–81.

85. Royle JA. N-mixture models for estimating population size from spatially replicated counts. *Biometrics*. 2004;60(1):108–15.
86. Royle JA, Dorazio RM. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data From Populations, Metapopulations and Communities*. Academic Press. 2008.
87. Royle JA, Link WA. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*. 2006;87(4):835–41.
88. Ruiz-Gutierrez V, Hooten MB, Grant Campbell EH. (in press). Uncertainty in surveillance and monitoring programs: a framework for citizen science data collection and analysis. *Methods in Ecology and Evolution*.
89. Russell J, Hanks E, Haran M. Dynamic models of animal movement with spatial point process interactions. *J Agric Biol Environ Stat*. 2016;21(1):22–40.
90. Sólymos P, Lele S, Bayne E. Conditional likelihood approach for analyzing single visit abundance survey data in the presence of zero inflation and detection error. *Environmetrics*. 2012;23(2):197–205.
91. Sólymos P, Lele SR. Revisiting resource selection probability functions and single-visit methods: clarification and extensions. *Methods Ecol Evol*. 2016;7(2):196–205.
92. Stoklosa J, Daly C, Foster SD, Ashcroft MB, Warton DI. A climate of uncertainty: accounting for error in climate variables for species distribution models. *Methods Ecol Evol*. 2015;6(4):412–23.
93. Thuiller W, Albert C, Araújo MB, Berry PM, Cabeza M, Guisan A, et al. Predicting global change impacts on plant species distributions: future challenges. *Perspectives in Plant Ecology Evolution and Systematics*. 2008;9(3):137–52.
94. Tye CA, McCleery RA, Fletcher RJ, Greene DU, Butryn RS. Evaluating citizen vs. professional data for modelling distributions of a rare squirrel. *J Appl Ecol*.
95. Tyre AJ, Tenhumberg B, Field SA, Niejalke D, Parris K, Possingham HP. Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecol Appl*. 2003;13(6):1790–1801.
96. VanDerWal J, Shoo LP, Graham C, Williams SE. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol Model*. 2009;220(4):589–94.
97. Warton DI, Blanchet FG, O’Hara RB, Ovaskainen O, Taskinen S, Walker SC, et al. So many variables: joint modeling in community ecology. *Trends Ecol Evol*. 2015;30(12):766–79.
98. Warton DI, Renner IW, Ramp D. Model-based control of observer bias for the analysis of presence-only data in ecology. *PLoS One*. 2013:e79168.
99. Warton DI, Shepherd LC. Poisson point process models solve the pseudo-absence problem for presence-only data in ecology. *Ann Appl Stat*. 2010;4(3):1383–1402.
100. Welsh AH, Lindenmayer DB, Donnelly CF. Fitting and interpreting occupancy models. *PLoS One*. 2013;8(1):e52015.
101. Wikle CK. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*. 2003;84(6):1382–94.
102. Wikle CK, Hooten MB. A general science-based framework for dynamical spatio-temporal models. *Test*. 2010;19(3):417–51.
103. Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant Campbell EH, et al. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol Evol*. 2013;4(3):236–43.
104. Zheng Y, Aukema BH. Hierarchical dynamic modeling of outbreaks of mountain pine beetle using partial differential equations. *Environmetrics*. 2010;21(7-8):801–16.