Statistical Reports

Ecology, 98(2), 2017, pp. 328–336
© 2016 by the Ecological Society of America

An integrated data model to estimate spatiotemporal occupancy, abundance, and colonization dynamics

Perry J. Williams,[1,2,8] Mevin B. Hooten,[2,3] Jamie N. Womble,[4,5] George G. Esslinger,[6] Michael R. Bower,[4] and Trevor J. Hefley[7]

[1]*Department of Fish, Wildlife, and Conservation Biology, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Department of Statistics, Colorado State University, Fort Collins, Colorado 80523 USA*
[3]*Department of Fish, Wildlife, and Conservation Biology, Colorado Cooperative Fish and Wildlife Research Unit, U.S. Geological Survey, Colorado State University, Fort Collins, Colorado 80523 USA*
[4]*Southeast Alaska Inventory and Monitoring Network, National Park Service, 3100 National Park Rd, Juneau, Alaska 99801 USA*
[5]*Glacier Bay Field Station, National Park Service, 3100 National Park Rd, Juneau, Alaska 99801 USA*
[6]*Alaska Science Center, U.S. Geological Survey, 4210 University Drive, Anchorage, Alaska 99508 USA*
[7]*Department of Statistics, Kansas State University, Manhattan, Kansas 66506 USA*

Abstract. Ecological invasions and colonizations occur dynamically through space and time. Estimating the distribution and abundance of colonizing species is critical for efficient management or conservation. We describe a statistical framework for simultaneously estimating spatiotemporal occupancy and abundance dynamics of a colonizing species. Our method accounts for several issues that are common when modeling spatiotemporal ecological data including multiple levels of detection probability, multiple data sources, and computational limitations that occur when making fine-scale inference over a large spatiotemporal domain. We apply the model to estimate the colonization dynamics of sea otters (*Enhydra lutris*) in Glacier Bay, in southeastern Alaska.

Key words: ecological diffusion; homogenization; hurdle model; integrated population model; partial differential equation; sea otters; state-space model; zero-inflated model.

Introduction

Characterizing species colonization/invasion (hereafter colonization) through space and time is a fundamental theme in ecology (Bullock et al. 2002). Applications include reintroduction of endangered species, invasive species management, and characterizing emerging/re-emerging wildlife or plant disease (Hooten et al. 2007, Hooten and Wikle 2008, Osnas et al. 2009, Marucco and McIntire 2010). Colonizations are dynamic processes, changing in space and time, and modeling these dynamics is imperative for ecological learning (Clark et al. 2001, Wikle 2003, Wikle and Hooten 2010). There has been a proliferation of statistical methods for modeling ecological dynamics (e.g., Wikle 2003, Wikle and Hooten 2006, 2010, Hooten et al. 2007, Cressie and Wikle 2011, Conn et al. 2015). The hierarchical statistical framework allows for the explicit incorporation of theoretical models that best represent our understanding of ecological systems (Hilborn and Mangel 1997, Cressie et al. 2009, Hobbs and Hooten 2015). Statistical models containing dynamic mechanisms for ecological colonization improve our ability to obtain accurate predictions in space and time while properly accommodating uncertainty pertaining to multiple sources of data, underlying mechanistic processes, and parameters (Wikle 2003, Hooten et al. 2007, Hobbs and Hooten 2015). Incorporating these dynamic models for ecological colonization improves our ability to obtain predictions in space and time, and include reliable measures of prediction error.

Two critical state-variables associated with species colonizations include the distribution of the species, and the abundance of the species within its distribution (i.e., occupancy and abundance, respectively; Brown 1984, 1995, Gaston 1996). There is a fundamental link between occupancy and abundance in both mathematics and ecological theory supported by empirical evidence (e.g., Gaston 1996, Welsh et al. 1996, Holt et al. 2002, Royle and Nichols 2003, Wenger and Freeman 2008, Smith et al. 2012). Explicitly estimating occupancy and

Manuscript received 26 May 2016; revised 2 October 2016; accepted 7 October 2016. Corresponding Editor: Donald R. Strong.
[8]E-mail: perry.williams@colostate.edu

abundance in a cohesive modeling framework has several advantages including: improved computational performance (Smith et al. 2012), straightforward model interpretation (Welsh et al. 1996, Smith et al. 2012), the ability to estimate additional latent parameters (Royle and Nichols 2003), and modeling rare or elusive species (Welsh et al. 1996). Although ecological models linking occupancy and abundance have been developed for static systems (e.g., Welsh et al. 1996, Smith et al. 2012), they have not been generalized for dynamic spatiotemporal processes: processes for which these models are most relevant. That is, explicitly incorporating occupancy and abundance in a spatiotemporal modeling framework will permit investigators to make combined inference or predictions of the dynamic colonization process, helping to identify future research and management priorities for species with expanding or contracting distributions.

We developed and implemented a spatiotemporal statistical framework that explicitly incorporates the mathematical link between occupancy and abundance and permits inference for each state variable through space and time. Our framework addresses several common challenges associated with analyzing spatiotemporal ecological data. First, we account for detection probability at two levels; the probability of detecting a species, given a sample site is occupied by the species, and the probability of detecting an individual after the site is determined to be occupied. Second, we demonstrate the inclusion of mechanistic process models (specifically, partial differential equations: PDE; Holmes et al. 1994) that incorporate ecological theory of population colonization, and perhaps more importantly, account for parameter uncertainty associated with the models. Third, because we developed the model within a Bayesian hierarchical framework, we demonstrate additional model flexibility, including the ability to combine multiple data sources to aid inference of model parameters, account for multiple levels of uncertainty, and account for nonlinear spatiotemporal dispersal and growth processes. Fourth, we discuss and implement a method for addressing computational limitations associated with modeling small-scale processes over large spatiotemporal domains.

To demonstrate the approach, we fit a spatiotemporal occupancy-abundance model for colonizations to three sources of aerial survey data collected on sea otter (*Enhydra lutris*) distribution and abundance in Glacier Bay, Alaska from 1993–2012. Sea otters in Glacier Bay were first detected at the mouth of Glacier Bay in 1993, and have since colonized the entire bay. We explicitly incorporate each data source to aid estimation of growth, diffusion, and detection probability.

Model Development

Following the terminology of Berliner (1996), we present a Bayesian hierarchical model consisting of three levels (Appendix S1). At the top level, we developed a *Data model* linking the observed data and associated variation to latent ecological processes. The data model we describe is specific to the combined inference on occupancy and abundance, although generalizations to other types of data (e.g., multi-state occupancy data) are straightforward within this framework. Next, we discuss *Process models* describing the underlying ecological processes (i.e., spatiotemporal occupancy and abundance dynamics). Finally, we developed *Parameter models* representing our prior knowledge about the parameter inputs in the ecological process model and data model. We use the square-bracket notation $[a|b]$ to represent the probability density function of variable a given variable b (Gelfand and Smith 1990), and also use $P(a = 1|b)$, to represent the probability of the realization $a = 1$, given b. We represent vectors with bold lowercase notation, and matrices with bold uppercase notation. Finally, we use parentheses to index processes that occur in continuous time and continuous space, and subscripts to index processes that occur in discrete time and discrete space.

Data model

Estimating species abundance often requires observers to first locate an occupied site (i.e., a sample unit with ≥1 individuals of a species of interest), and then count the number of individuals at the occupied site. Examples include locating a herd of ungulates, or a lek of grouse, and then counting the individuals within the heard or lek. Occupancy and abundance sampling is often conducted such that observers collect data over a disjoint set of known-area sites, as with a grid, where some sites are sampled and others are not. We let $y_{i,t}$ represent the ith observation ($i = 1, \ldots, k$) collected at a site indexed by the spatial location $\mathbf{s}_i$ (e.g., the center of the site) within a study area $\mathcal{S}$ ($\mathbf{s}_i \in \mathcal{S} \subset \mathcal{R}^2$), during time $t = 1, \ldots, T$, in which $y_{i,t} = 1$ if ≥1 individual was detected at the site and $y_{i,t} = 0$ if no individual was detected. We assume the site indexed by $\mathbf{s}_i$ is a discrete grid cell matching the observation process in which ≥1 individual could be detected. Subsequent to detecting ≥1 individual at a site, observers have the opportunity to count the number of individuals at the site. Let $c_{i,t}$ represent count data with support in the non-negative integers. The data $c_{i,t}$ often underestimate true abundance $n_{i,t}$ due to imperfect detection probability. Therefore, it is necessary to account for detection probability (p) to estimate true abundance. We represent the relationship among count data, true abundance, and observed occupancy data as a hurdle model

$$c_{i,t} \sim \begin{cases} [c_{i,t}|n_{i,t},p], & y_{i,t} = 1 \\ 0, & y_{i,t} = 0 \end{cases} \tag{1}$$

where $[c_{i,t}|n_{i,t}, p]$ is a probability mass function relating the counts $c_{i,t}$ to the latent, true abundance $n_{i,t}$ and detection probability p (e.g., the zero-truncated (ZT) binomial distribution). Note that, in Eq. 1, $P(c_{i,t} = 0|y_{i,t} = 1) = 0$, and $P(c_{i,t} > 0|y_{i,t} = 0) = 0$. That is, if the species was detected, at least one animal was counted; if the species was not detected, it could not be counted.

Species at occupied sites are not always detected and, to estimate true occupancy status in these cases, it is necessary to estimate the species-level detection probability ρ. We represent the relationship between observed occupancy data $y_{i,t}$ and true occupancy as a zero-inflated model

$$y_{i,t} \sim \begin{cases} [y_{i,t}|\rho], & n_{i,t} > 0 \\ 0, & n_{i,t} = 0 \end{cases} \quad (2)$$

where $[y_{i,t}|\rho]$ is a probability mass function conditioning occupancy data $y_{i,t}$ on the latent, true abundance $n_{i,t}$ and species-level detection probability ρ (e.g., the Bernoulli distribution). Both p and ρ may depend on spatial or temporal covariates related to the observation process. Additionally, ρ is likely to be influenced by the abundance of the species at a site, $n_{i,t}$ (i.e., species-level detection probability is likely positively related to abundance). For example, if each individual within an occupied area has an equal detection probability (r), and there is independence of detections among individuals, then

$$\rho_{i,t} = 1 - (1 - r)^{n_{i,t}} \quad (3)$$

(Royle and Nichols 2003). There is a subtle distinction between r and p, although both relate to the probability of detecting an individual. The parameter r is the probability of detecting an individual in an occupied site for which the observer does not know the site is occupied (i.e., the unconditional individual detection probability). The parameter p is the probability of detecting an individual within an occupied site for which the observer knows the site is occupied (i.e., the individual detection probability, conditional on observing occupancy). If the process of observing an individual is the same, regardless of knowledge of the occupancy status, then $r = p$. If the process is not the same (e.g., investigators search more diligently for individuals in sites known to be occupied or vice versa), then r and p could be estimated separately. If all individuals in an occupied site do not have the same detection probability r, alternative functions linking abundance to the probability of detecting the species can be easily incorporated. For example, $\text{logit}(\boldsymbol{\rho}) = I_{\{n>0\}} \times (b_0 + b_1\mathbf{x}_1 + \cdots + b_l\mathbf{x}_l)$, for l covariates that affect detection probability, and where $I_{\{n>0\}}$ is an indicator variable that equals one when $n > 0$, and zero otherwise.

Process model

Abundance.—An overview of spatiotemporal models to estimate animal abundance was provided by Conn et al. (2015). We describe a specific process model for abundance that is sufficiently flexible to address many ecological questions. We use a Poisson distribution to model the latent, true abundance $n_{i,t}$, with a dynamically evolving mean and variance parameter $\lambda(\mathbf{s}_i, t)$. That is, $n_{i,t} \sim \text{Poisson}(\lambda(\mathbf{s}_i, t))$, where $\lambda(\mathbf{s}_i, t)$ is the abundance intensity, a spatiotemporal dynamic process. In general, $\lambda(\mathbf{s}_i, t) = \int_{B_i} \lambda(\mathbf{u}, t) d\mathbf{u}$, where B_i is the unit in which counts occur. In our sea otter example, we assume the scale at which data were collected coincides with the numerical scale in which we solved $\lambda(\mathbf{u}, t)$, precluding the necessity to integrate over B_i. If data at a site are collected from a larger spatial area than that used to numerically solve $\lambda(\mathbf{u}, t)$, then the integration would be necessary. Assuming Markovian dynamics, we can express the abundance intensity for time $t \in \mathcal{T}$ as a function of the previous abundance intensity and model parameters that describe the dynamics of the process $\lambda(\mathbf{s}, t) = h(\lambda(\mathbf{s}, t - \Delta t); \theta)$ (Wikle and Hooten 2006, 2010). The function h is general and can be chosen to match the specifics of the study. We used the PDE known as *ecological diffusion* to describe diffusion and growth dynamics (Garlick et al. 2011, Hooten et al. 2013), because of its explicit connection to modeling animal movement using abundance data (c.f., Hooten et al. 2010, Hooten and Wikle 2010, for modeling animal movement using individuals directly). The ecological diffusion PDE, in two dimensions with Malthusian growth, is

$$\frac{\partial \lambda(\mathbf{s}, t)}{\partial t} = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} \right) \delta(\mathbf{s}, t)\lambda(\mathbf{s}, t) + \gamma(\mathbf{s}, t)\lambda(\mathbf{s}, t) \quad (4)$$

where $\partial\lambda(\mathbf{s}, t)/\partial t$ represents the instantaneous change in abundance intensity over the continuous spatial domain $\mathbf{s} \equiv (s_1, s_2)' \in \mathcal{S}$ during time t, $(\partial^2/\partial s_1^2 + \partial^2/\partial s_2^2)$ is the differential operator, $\delta(\mathbf{s}, t)$ represents motility, and $\gamma(\mathbf{s}, t)$ represents the instantaneous growth rate. Both $\delta(\mathbf{s}, t)$ and $\gamma(\mathbf{s}, t)$ can depend on covariates that vary over space and time, although we focus on space. We consider a log-linear relationship between the motility $\delta(\mathbf{s}, t)$ and covariates, and a linear relationship between $\gamma(\mathbf{s}, t)$ and covariates. Specifically, $\log(\boldsymbol{\delta}) = \mathbf{X}\boldsymbol{\beta}$, and $\boldsymbol{\gamma} = \mathbf{W}\boldsymbol{\alpha}$, where $\mathbf{X}$ is a matrix of covariates thought to influence motility, $\mathbf{W}$ is a matrix of covariates thought to influence growth rate, and $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are parameters to be estimated. Although we focus on Malthusian growth $\gamma(\mathbf{s})\lambda(\mathbf{s}, t)$, other continuous growth functions are possible (e.g., Holmes et al. 1994, Turchin 2003).

Calculating the abundance intensity $\lambda(\mathbf{s}, t)$ requires solving Eq. 4 to describe how $\lambda(\mathbf{s}, t)$ changes through space and time. We use finite-differencing to solve Eq. 4, and subsequently calculate $\lambda(\mathbf{s}, t)$ (Hooten et al. 2013). Finite differencing consists of partitioning the continuous spatial domain into q_s grid cells and partitioning the continuous temporal domain into q_t bins of width Δt. The discrete partitioning results in a linear equation for calculating $\lambda(\mathbf{s}, t)$ based on $\lambda(\mathbf{s}, t - \Delta t)$. That is, $\lambda_t = \mathbf{H}\lambda_{t-\Delta t}$, where $\mathbf{H}$ is a $q_s \times q_s$ transition (or propagator) matrix that describes how $\lambda(\mathbf{s}, t)$ changes in each discrete time step, as a function of other cells in the spatial domain, and the growth and diffusion parameters $\gamma(\mathbf{s})$ and $\delta(\mathbf{s})$. The $\mathbf{H}$ matrix for the discretized ecological diffusion model, using first-order finite differencing, is five-diagonal, representing that $\lambda(\mathbf{s}_i, t)$ depends

on five cells in time $t - \Delta t$, cell i and its four closest neighbors. Calculation of $\mathbf{H}$ based on $\delta(\mathbf{s})$ and $\gamma(\mathbf{s})$ is straightforward and we provide R code for this calculation in Appendix S2.

For initial conditions of the PDE, we selected a function that allowed us to incorporate information on both the intensity and dispersion of $\lambda(\mathbf{s}, 0)$, based on an epicenter where the invasion was initiated. A scaled Gaussian kernel is a standard function that meets these conditions and is described as

$$\lambda(\mathbf{s},0) = \frac{\theta e^{\frac{-|\mathbf{s}-\mathbf{d}|^2}{\kappa^2}}}{\int_S e^{\frac{-|\mathbf{s}-\mathbf{d}|^2}{\kappa^2}} \mathrm{d}\mathbf{s}}$$

where θ is a scale parameter controlling the initial density, κ is a dispersion parameter controlling the radial distance of the initial density, and $\mathbf{d}$ is the epicenter. We chose a reflective spatial boundary condition (sensu Cantrell and Cosner 2004) for $\lambda(\mathbf{s}, t)$ at spatial locations adjacent to terrestrial environments, because sea otters are typically restricted to marine environments. A reflective boundary is relevant for aquatic species living in a body of water or other systems when there exists a barrier that the individual cannot cross (e.g., sea otters in Glacier Bay, which is surrounded by land except at the mouth of the bay). We did not use a reflective boundary condition at the mouth of Glacier Bay, allowing diffusion to occur in and out of the bay. Including boundaries, the linear equation for abundance intensity can be written as $\boldsymbol{\lambda}_t = \mathbf{H}\boldsymbol{\lambda}_{t-\Delta t} + \mathbf{H}^{(b)}\boldsymbol{\lambda}^{(b)}_{t-\Delta t}$.

Occupancy.—The link to occupancy from abundance can be derived from the rules of probability. Assuming $n_{i,t} \sim \text{Poisson}(\lambda(\mathbf{s}_i, t))$, then the occupancy probability is $\varphi_{i,t} = P(n_{i,t} > 0|\lambda(\mathbf{s}_i, t)) = 1 - e^{-\lambda(\mathbf{s}_i,t)}$. Assuming $y_{i,t}|n_{i,t} \sim \text{Bernoulli}(1-(1-r)^{n_{i,t}})$ (as in Eqs. 2 and 3); then, marginalizing the Bernoulli–Poisson mixture yields the following probabilities of non-detection and detection of the species, respectively: $P(y_{i,t} = 0) = e^{-r\lambda(\mathbf{s}_i,t)}$ and $P(y_{i,t} = 1) = 1 - e^{-r\lambda(\mathbf{s}_i,t)}$.

Parameter models

To complete the Bayesian specification of the spatiotemporal occupancy-abundance model, we describe probability models for the parameters discussed in the data and process components, above. Parameters requiring prior distributions include $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, p, ρ, θ, and κ. Assuming Malthusian growth in Eq. 4, $\gamma(\mathbf{s})$ represents the intrinsic rate of population growth, and should be parameterized accordingly. The parameters p and ρ represent individual and species-level detection probabilities and should be restricted between 0 and 1. Note that, if ρ is deterministically linked to n and r, as in Eq. 3, no parameter model is needed for ρ. However, if $r \neq p$, a parameter model is needed for r. Finally, θ and κ must be positive. We provide specific values for these parameters in our application to sea otter data in Appendix S1.

The full Bayesian posterior distribution is

$$[\mathbf{N},\boldsymbol{\Lambda},p,r,\boldsymbol{\alpha},\boldsymbol{\beta},\theta,\kappa|\mathbf{C},\mathbf{Y}] \propto \prod_{t=1}^{T}\prod_{i=1}^{k}\left\{[c_{i,t}|n_{i,t},p]^{y_{i,t}} I^{1-y_{i,t}}_{\{c_{i,t}=0\}}[y_{i,t}|r,n_{i,t}][n_{i,t}|\theta,\kappa,\boldsymbol{\alpha},\boldsymbol{\beta}]\right\} \times [p][r][\theta][\kappa][\boldsymbol{\alpha}][\boldsymbol{\beta}] \quad (5)$$

where $I^{1-y_{i,t}}_{\{c_{i,t}=0\}}$ is an indicator variable for sites where no species were detected, and therefore $c_{i,t} = 0$ with probability 1.

Computational efficiency

When modeling spatiotemporal dynamics at a fine scale (i.e., q_s and q_t are large), computation time can be prohibitive because the matrix multiplication to evaluate $\lambda_t = \mathbf{H}_t\lambda_{t-\Delta t}$ results in computation cost that scales by a factor of $q_s^2 q_t$. As a result, numerical methods that efficiently solve PDEs for implementation within statistical models are required for large spatial and temporal domains. Homogenization is a method that facilitates solving PDEs at a course scale, reducing q_s, then optimally down-scales the results to a fine scale (Garlick et al. 2011, Hooten et al. 2013). We use homogenization to fit the model in the sea otter example below. Additionally, we provided R code to demonstrate homogenization in Appendix S2.

Sea Otter Occupancy and Abundance

We fit the spatiotemporal occupancy-abundance model to three sources of aerial survey data collected on sea otters in Glacier Bay, Alaska (Appendix S3). The first data source was obtained using a design-based survey where observers flew in an airplane along transects that were 400 m wide, and systematically placed across Glacier Bay during 1999–2004, 2006, and 2012 (Appendix S4; Bodkin and Udevitz 1999). We discretized the continuous spatial domain into 400 × 400 m grid cells to match the data collection process. During surveys, observers determined the presence or absence of ≥1 otters within a 400 × 400 m site, and subsequently counted individual otters within occupied sites. Thus, the data included presence or absence ($y_{i,t}$) and counts ($c_{i,t}$) of otters at a site. The second data source included 469 grid cells/sites comprising a randomly selected subset of the design-based survey. At these 469 sites, after ≥1 otters were detected and counted using the same procedures from the design-based survey, five concentric circles were flown around the site so observers could obtain precise counts of abundance. The concentric circles were flown in 3.7 min, a time chosen based on the aerobic dive limit of sea otters (so diving otters could be included in the counts when they resurfaced). Thus, these data include information on $n_{i,t}$ at the site, for comparison to $c_{i,t}$, to estimate p. The third data source consisted of another aerial survey intended to estimate the spatial distribution of sea otters. The third data set was acquired in 1993,

1996, 1997, 1998, 2005, 2009, and 2010. These data contained presence ($y_{i,t}$ = 1) and count data ($c_{i,t}$ > 0) only, and no absence data. We used these *distributional data* to supplement the design-based data in years when the design-based survey was not conducted. Because our model formulation does not require presence and absence data to fit the model, both the design-based data and the distributional data could be used. The additional absence information from the design-based data allows improved precision when estimating model parameters, while missing absence data from the distributional surveys had no effect on parameter estimates.

To fit the model to these data, we assumed a ZT-binomial distribution for $c_{i,t}$ in Eq. 1, and a Bernoulli distribution for $y_{i,t}$ in Eq. 2. We used Eq. 3 to model group-level detection probability, assuming $r = p$. Given the model specification, setting $r = p$ improved computation speed due to the resulting conjugacy when sampling n (Appendix S2). We assumed $n_{i,t} \sim \text{Poisson}(\lambda(\mathbf{s}_i, t))$. For our ecological diffusion model, we assumed motility was exponentially related to depth (either <40 m or not), distance to shore, slope of the ocean floor, and shoreline complexity. That is, $\log(\delta(\mathbf{s}_i)) = \beta_0 + \beta_1 \text{depth}_i + \beta_2 \text{dist}_i + \beta_3 \text{slope}_i + \beta_4 \text{complexity}_i$. We included an intercept, and no covariates for growth, so that $\gamma = \alpha$. We selected an epicenter **d** outside the mouth of Glacier Bay, where sea otters were observed prior to their colonization of Glacier Bay. We used vague prior distributions for all parameters (provided in Appendix S1).

The fine-scale spatial resolution of grid cells was originally 400 × 400 m, which matched the width of aerial transects. We up-scaled the spatial resolution by a factor of 100 (4,000 × 4,000 m) using homogenization. To fit the model, we discretized the temporal domain into $\Delta t \approx 4$ d. We obtained summaries of the posterior distribution using an MCMC algorithm in R version 3.0.2 (R Core Team 2013) and C++. We calculated marginal posterior distributions for the parameters: $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, p, θ, and κ, and the derived parameters: $\boldsymbol{\Lambda}$ (abundance intensity), **N** (abundance), $\boldsymbol{\Phi}$ (occupancy probability), and **P** (species-level detection probability), each a $q_s \times T$ matrix. We obtained three parallel chains of 100,000 iterations using our algorithm, discarding the first

TABLE 1. Posterior mean and 95% credible intervals for estimated parameters in the sea otter occupancy-abundance model.

Parameter	Lower bound	Mean	Upper bound
$\sum_{i=1}^{q_s} n_{i,1993}$ (abundance in 1993)	588	637	687
$\sum_{i=1}^{q_s} n_{i,1994}$	654	693	733
$\sum_{i=1}^{q_s} n_{i,1995}$	669	705	742
$\sum_{i=1}^{q_s} n_{i,1996}$	718	754	789
$\sum_{i=1}^{q_s} n_{i,1997}$	795	831	868
$\sum_{i=1}^{q_s} n_{i,1998}$	896	934	972
$\sum_{i=1}^{q_s} n_{i,1999}$	1,019	1,058	1,098
$\sum_{i=1}^{q_s} n_{i,2000}$	1,164	1,205	1,247
$\sum_{i=1}^{q_s} n_{i,2001}$	1,331	1,376	1,420
$\sum_{i=1}^{q_s} n_{i,2002}$	1,523	1,570	1,618
$\sum_{i=1}^{q_s} n_{i,2003}$	1,741	1,792	1,843
$\sum_{i=1}^{q_s} n_{i,2004}$	1,988	2,044	2,101
$\sum_{i=1}^{q_s} n_{i,2005}$	2,267	2,330	2,394
$\sum_{i=1}^{q_s} n_{i,2006}$	2,581	2,653	2,727
$\sum_{i=1}^{q_s} n_{i,2007}$	2,933	3,020	3,108
$\sum_{i=1}^{q_s} n_{i,2008}$	3,330	3,436	3,543
$\sum_{i=1}^{q_s} n_{i,2009}$	3,777	3,907	4,039
$\sum_{i=1}^{q_s} n_{i,2010}$	4,282	4,442	4,605
$\sum_{i=1}^{q_s} n_{i,2011}$	4,850	5,048	5,250
$\sum_{i=1}^{q_s} n_{i,2012}$	5,491	5,735	5,985
β_0 (intercept)	17.72	17.78	17.84
β_1 (depth)	−1.48	−1.42	−1.36
β_2 (distance to shore)	0.77	0.81	0.85
β_3 (bottom slope)	−0.35	−0.33	−0.31
β_4 (shoreline complexity)	1.00	1.03	1.07
γ (intrinsic growth rate)	0.20	0.21	0.22
θ (initial condition for density)	1910	2058	2213
κ (initial condition for dispersal)	5.95	6.59	7.28
p (individual detection probability)	0.74	0.76	0.78

Note: Monte Carlo standard error was < 3 for θ and $\sum_{i=1}^{q_s} n_{i,t}$ and < 0.01 for all other parameters, where $n_{\{i,t\}}$ is the latent sea otter abundance at site i = 1, …, $q_{\{s\}}$ during time t = 1993, …, 2012.

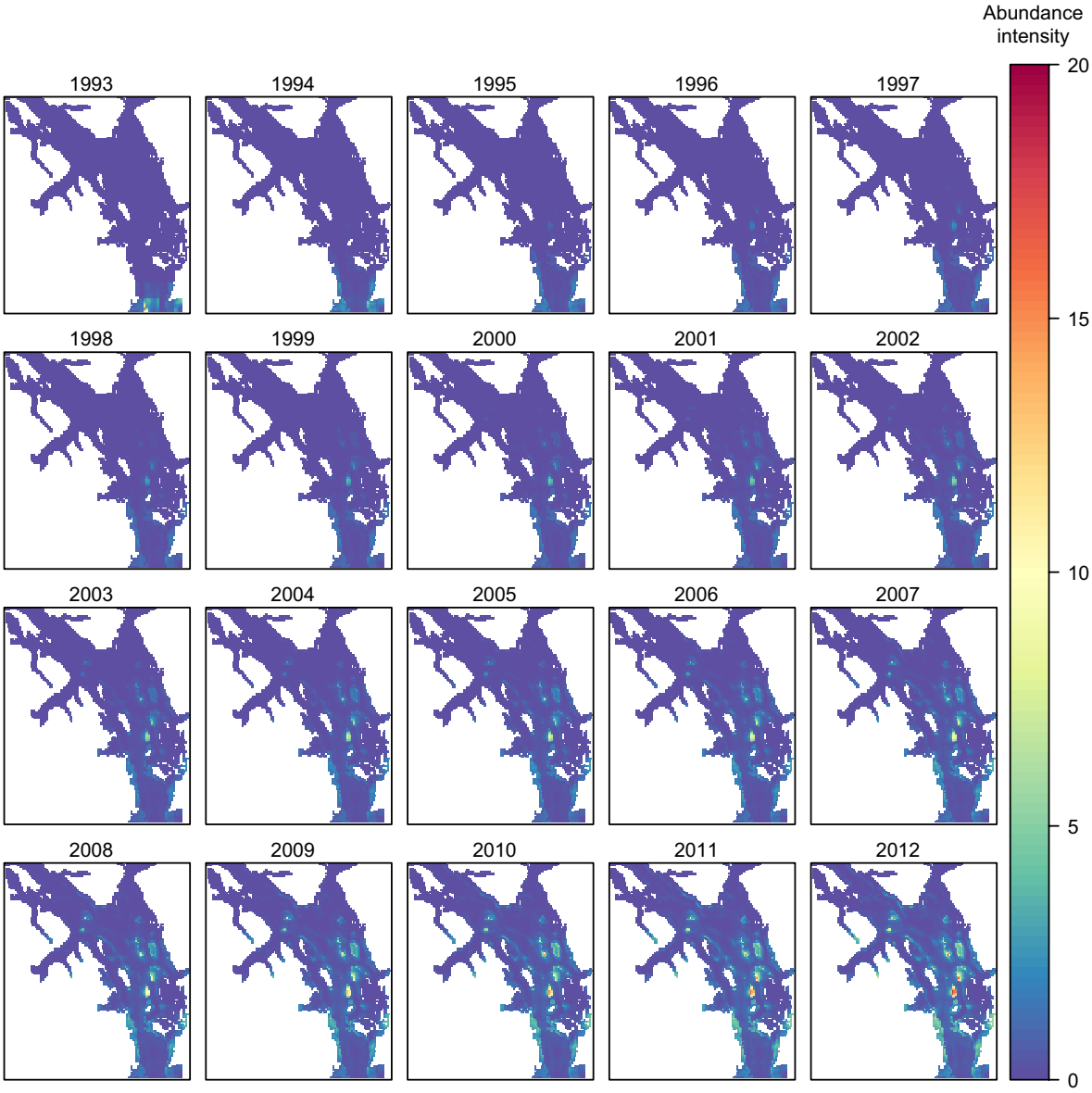

FIG. 1. Estimated posterior mean (also variance, due to the mean--variance relationship of the Poisson distribution) of abundance intensity, $\lambda(\mathbf{s}, t)$, of sea otters in Glacier Bay, Alaska from 1993–2012.

50,000 burn-in iterations. We examined convergence using trace plots and Gelman-Rubin diagnostics. We estimated Monte Carlo error for all estimates (Brooks et al. 2011, Dorazio 2016; Table 1). The full model statement and MCMC algorithm are provided in Appendix S1 and Appendix S2, respectively.

We report means of posterior distributions, which are optimal for estimation under squared-error loss (e.g., Bernardo and Juárez 2003, Williams and Hooten 2016). Posterior abundance estimates for sea otters in Glacier Bay are shown in Fig. 1 and Table 1. Posterior mean occupancy probabilities are shown in Fig. 2. Additional posterior parameter means and credible intervals are shown in Table 1. Our ecological diffusion model accounts for variation in motility due to habitat covariates, and predicts that animals will eventually accumulate in desirable habitats and avoid, or move quickly through undesirable habitats (Garlick et al. 2011). Our diffusion results suggest sea otters were influenced by habitat type, and sea otters accumulated in shallow areas, close to shore, with steep bottom slope and relatively little shoreline complexity. They diffused quickly through or avoided areas not having some or all of these traits. The posterior mean intrinsic rate of growth was 0.21, which was similar to the estimated maximum reproductive rate for sea otters (Estes 1990). The posterior mean probability of detecting an individual sea otter was 0.76. Posterior mean species-level detection probability depended on abundance, n, and is

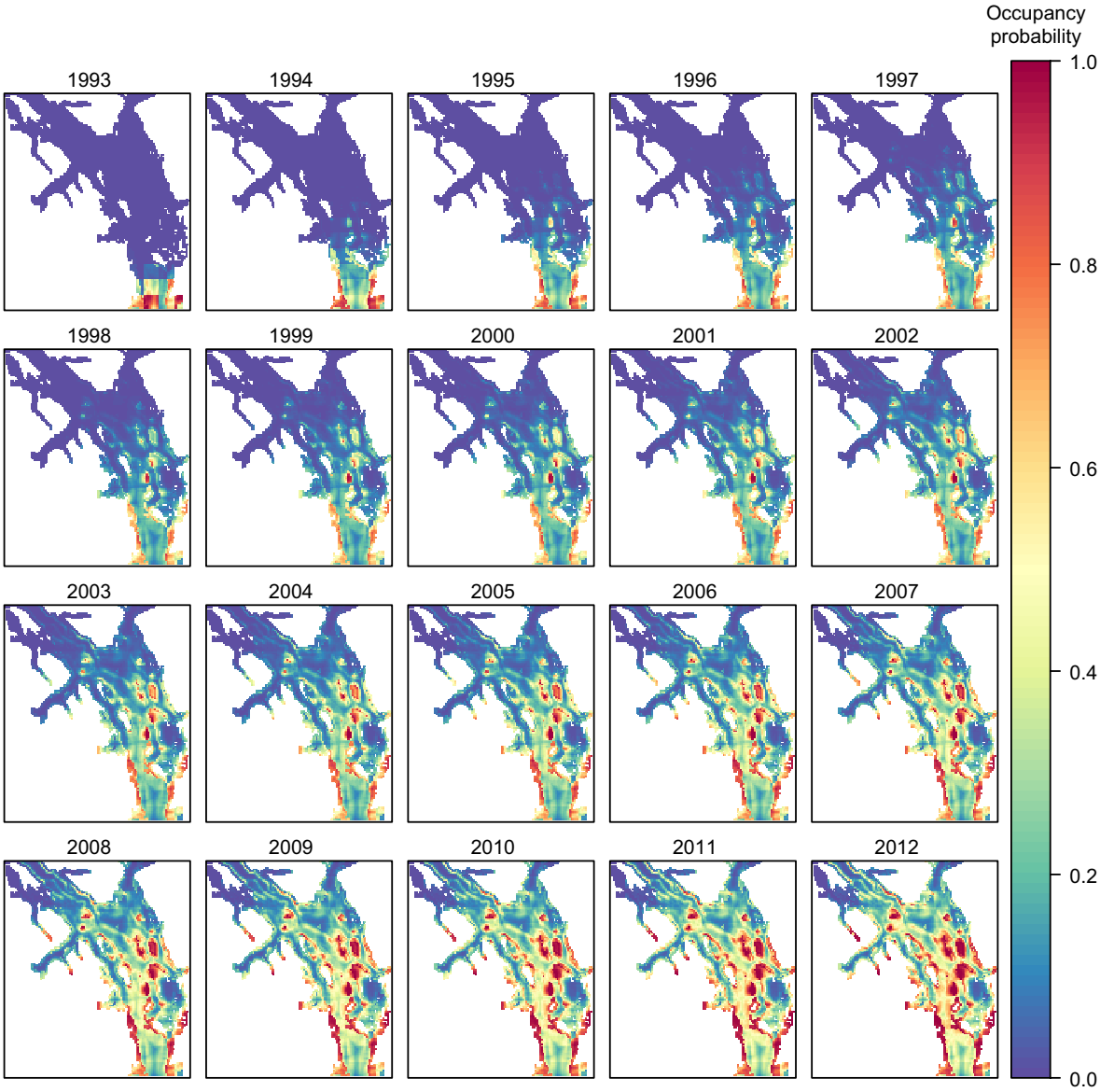

FIG. 2. Estimated posterior mean occupancy probability, $\varphi_{i,t}$, of sea otters in Glacier Bay, Alaska from 1993–2012.

calculated as $\rho = 1 - 0.24^n$ (e.g., for $n = 2, 3, 4$, $\rho = 0.94, 0.99, 1.0$, respectively).

DISCUSSION

Many surveys in ecology follow the procedure of first locating an area occupied by a species, then expending additional effort to count the number of individuals in the occupied area. For example, counting amphibian egg masses at ponds where the species were heard calling, counting individual moose after a group has been detected, or counting the number of individuals present after a lek is observed (Steinhorst and Samuel 1989, McDonald et al. 2014). Design-based cluster sampling was developed for problems of a similar nature (Thompson 1990). Several statistical models have been developed that incorporate the functional relationship between occupancy and abundance (e.g., Welsh et al. 1996, Royle and Nichols 2003, Wenger and Freeman 2008, Smith et al. 2012). We proposed a spatiotemporal framework that generalizes previous research on incorporating the intraspecific relationship between occupancy and abundance, facilitates the estimation of detection probability at multiple levels, and is straightforward to estimate and interpret.

Although we focus on abundance, the formulation in Eq. 1 is sufficiently general to include other state variables instead of $n_{i,t}$. For example, Nichols et al. (2007) developed a modeling framework for estimating occupancy with multiple states and state uncertainty (i.e., a multi-state

occupancy model). The multi-state occupancy model is a specific form of Eq. 1. That is, assume investigators are interested in the probability of occupancy, and the probability of reproduction at an occupied site (as in Nichols et al. 2007). By parameterizing Eq. 1, such that $[c_{i,t}|n_{i,t}, p_{i,t}]$ is a Bernoulli distribution, $n_{i,t}$ represents breeding success, $c_{i,t}$ represents whether breeding was observed, and $p_{i,t}$ represents the probability of observing breeding success given it occurred, we arrive at a multi-state occupancy model analogous to that described by Nichols et al. (2007). In addition to multi-state occupancy models, other state variables could be considered by appropriately modeling $[c_{i,t}|n_{i,t}, p_{i,t}]$ (e.g., using the multinomial distribution common in capture–recapture studies).

We demonstrated the implementation of the model using ecological diffusion. Ecological diffusion is flexible and applies to many ecological systems. Although we used ecological diffusion, any process model reflecting ecological theory can be easily incorporated into an occupancy–abundance framework (e.g., Broms et al. 2016). Finally, to solve the ecological diffusion PDE efficiently, we relied on the multi-scale analysis technique, homogenization. Homogenization may be useful for many spatiotemporal analyses in which fine-scale inference is desired over a large spatiotemporal domain.

Acknowledgments

The authors thank Paul Conn, Jacob Ivan, and two anonymous reviewers for valuable insight about this work. Funding was provided from the National Park Service Inventory and Monitoring Program and NSF DMS 1614392. Heather Coletti, Dan Esler, and Dan Monson provided technical and logistical support. Data were collected by James Bodkin, Janet Doherty, Dan Monson, and Ben Weitzman. Pat Kearney and Andy Harcombe piloted survey airplanes. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Berliner, L. M. 1996. Hierarchical Bayesian time series models. Pages 15–22 *in* K. M. Hanson and R. N. Silver, editors. Maximum entropy and Bayesian methods. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Bernardo, J. M., and M. Juárez. 2003. Intrinsic estimation. Bayesian Statistics 7:465–476.

Bodkin, J. L., and M. S. Udevitz. 1999. An aerial survey method to estimate sea otter abundance. Pages 13–26 *in* G. W. Garner, S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. McDonald, and D. G. Robertson, editors. Marine mammal survey and assessment methods. A. A. Balkema, Leiden, The Netherlands.

Broms, K. M., M. B. Hooten, D. S. Johnson, R. Altwegg, and L. L. Conquest. 2016. Dynamic occupancy models for explicit colonization processes. Ecology 97:194–204.

Brooks, S., A. Gelman, G. Jones, and X.-L. Meng. 2011. Handbook of Markov chain Monte Carlo. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Brown, J. H. 1984. On the relationship between abundance and distribution of species. American Naturalist 124:255–279.

Brown, J. H. 1995. Macroecology. University of Chicago Press, Chicago, Illinois, USA.

Bullock, J. M., R. E. Kenward, and R. S. Hails. 2002. Dispersal ecology: 42nd Symposium of the British Ecological Society. Cambridge University Press, Cambridge, UK.

Cantrell, R. S., and C. Cosner. 2004. Spatial ecology via reaction-diffusion equations. John Wiley & Sons, Chichester, UK.

Clark, J. S., et al. 2001. Ecological forecasts: an emerging imperative. Science 293:657–660.

Conn, P. B., D. S. Johnson, J. M. V. Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235–252.

Cressie, N., and C. K. Wikle. 2011. Statistics for spatio-temporal data. John Wiley & Sons, Hoboken, New Jersey, USA.

Cressie, N., C. A. Calder, J. S. Clark, J. M. V. Hoef, and C. K. Wikle. 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecological Applications 19:553–570.

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

Estes, J. A. 1990. Growth and equilibrium in sea otter populations. Journal of Animal Ecology 59:385–401.

Garlick, M. J., J. A. Powell, M. B. Hooten, and L. R. McFarlane. 2011. Homogenization of large-scale movement models in ecology. Bulletin of Mathematical Biology 73:2088–2108.

Gaston, K. J. 1996. The multiple forms of the interspecific abundance-distribution relationship. Oikos 76:211–220.

Gelfand, A. E., and A. F. Smith. 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85:398–409.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, New Jersey, USA.

Hobbs, N. T., and M. B. Hooten. 2015. Bayesian models: a statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.

Holmes, E. E., M. A. Lewis, J. Banks, and R. Veit. 1994. Partial differential equations in ecology: spatial interactions and population dynamics. Ecology 75:17–29.

Holt, A. R., K. J. Gaston, and F. He. 2002. Occupancy-abundance relationships and spatial distribution: a review. Basic and Applied Ecology 3:1–13.

Hooten, M. B., and C. K. Wikle. 2008. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. Environmental and Ecological Statistics 15:59–70.

Hooten, M. B., and C. K. Wikle. 2010. Statistical agent-based models for discrete spatio-temporal systems. Journal of the American Statistical Association 105:236–248.

Hooten, M. B., C. K. Wikle, R. M. Dorazio, and J. A. Royle. 2007. Hierarchical spatiotemporal matrix models for characterizing invasions. Biometrics 63:558–567.

Hooten, M. B., D. S. Johnson, E. M. Hanks, and J. H. Lowry. 2010. Agent-based inference for animal movement and selection. Journal of Agricultural, Biological, and Environmental Statistics 15:523–538.

Hooten, M. B., M. J. Garlick, and J. A. Powell. 2013. Computationally efficient statistical differential equation modeling using homogenization. Journal of Agricultural, Biological and Environmental Statistics 18:405–428.

Marucco, F., and E. McIntire. 2010. Predicting spatio-temporal recolonization of large carnivore populations and livestock depredation risk: wolves in the Italian Alps. Journal of Applied Ecology 47:789–798.

McDonald, L., et al. 2014. Range-wide population size of the lesser prairie-chicken: 2012 and 2013. Wildlife Society Bulletin 38:536–546.

Nichols, J. D., J. E. Hines, D. I. Mackenzie, M. E. Seamans, and R. Gutierrez. 2007. Occupancy estimation and modeling with multiple states and state uncertainty. Ecology 88:1395–1400.

Osnas, E. E., D. M. Heisey, R. E. Rolley, and M. D. Samuel. 2009. Spatial and temporal patterns of chronic wasting disease: fine-scale mapping of a wildlife epidemic in Wisconsin. Ecological Applications 19:1311–1322.

R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence–absence data or point counts. Ecology 84:777–790.

Smith, A. N., M. J. Anderson, and R. B. Millar. 2012. Incorporating the intraspecific occupancy–abundance relationship into zero-inflated models. Ecology 93:2526–2532.

Steinhorst, R. K., and M. D. Samuel. 1989. Sightability adjustment methods for aerial surveys of wildlife populations. Biometrics 45:415–425.

Thompson, S. K. 1990. Adaptive cluster sampling. Journal of the American Statistical Association 85:1050–1059.

Turchin, P. 2003. Complex population dynamics: a theoretical/empirical synthesis. Princeton University Press, Princeton, New Jersey, USA.

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

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

Wikle, C. K. 2003. Hierarchical Bayesian models for predicting the spread of ecological processes. Ecology 84:1382–1394.

Wikle, C. K., and M. B. Hooten. 2006. Hierarchical Bayesian spatio-temporal models for population spread. Pages 145–169 *in* J. S. Clark and A. Gelfand, editors. Hierarchical modelling for the environmental sciences. Oxford University Press, Oxford, UK.

Wikle, C. K., and M. B. Hooten. 2010. A general science-based framework for dynamical spatio-temporal models. Test 19:417–451.

Williams, P. J., and M. B. Hooten. 2016. Combining statistical inference and decisions in ecology. Ecological Applications 26:1930–1942.

Supporting Information

Additional supporting information may be found in the online version of this article at http://onlinelibrary.wiley.com/doi/10.1002/ecy.1643/suppinfo