

# Estimating the Use of Public Lands: Integrated Modeling of Open Populations with Convolution Likelihood Ecological Abundance Regression

Lutz F. Gruber<sup>\*</sup>, Erica F. Stuber<sup>†</sup>, Lyndsie S. Wszola<sup>‡</sup>, and Joseph J. Fontaine<sup>§</sup>

**Abstract.** We present an integrated open population model where the population dynamics are defined by a differential equation, and the related statistical model utilizes a Poisson binomial convolution likelihood. Key advantages of the proposed approach over existing open population models include the flexibility to predict related, but unobserved quantities such as total immigration or emigration over a specified time period, and more computationally efficient posterior simulation by elimination of the need to explicitly simulate latent immigration and emigration. The viability of the proposed method is shown in an in-depth analysis of outdoor recreation participation on public lands, where the surveyed populations changed rapidly and demographic population closure cannot be assumed even within a single day.

**Keywords:** integrated population model, panel data methods, discrete valued distributions, unevenly spaced time series.

## 1 Introduction

Ever-increasing pressures on ecological systems necessitate effective conservation and management plans for animal and plant populations. Accurate population models that account for social, economic, and ecological processes driving population dynamics, including birth, survival, and movement, can help predict population responses to natural and anthropogenic change. More reliable population projections can help to facilitate protecting or recovering endangered populations (Hanski and Simberloff, 1997), the development of effective public health strategies (Brauer and Castillo-Chavez, 2001), and improved understanding of geographical mobility and socio-economics (Schoen, 1988).

Our ability to detect changes in populations is constrained by the spatial and temporal resolution of data collected. The spatial and temporal resolution of data on species' abundances and their distribution is limited by the high cost of data collection, often in-

---

<sup>\*</sup>Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA, [lutz.gruber@gmail.com](mailto:lutz.gruber@gmail.com)

<sup>†</sup>Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA

<sup>‡</sup>Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA

<sup>§</sup>U.S. Geological Survey Nebraska Cooperative Fish and Wildlife Research Unit, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, USA

ducing a trade-off between spatial and temporal resolution of data collection. Although traditionally difficult to obtain, high temporal-resolution population estimates are of particular interest for social and ecological inference, especially for populations with fast temporal dynamics or populations experiencing rapid environmental change (Sih et al., 2011). Increasingly accessible and inexpensive technology such as camera traps and remote sensing is empowering researchers to collect data with finer temporal resolution at large spatial scales. For example, while many ecological studies have traditionally collected data from a given location a few times per year, and assume demographic population closure (that is, no individuals added or removed during a set time period) (for example, Stuber et al., 2018; Huber et al., 2017; Kéry and Schmid, 2006; Chandler et al., 2015; Morin et al., 2017), time-lapse photography facilitates multiple sample collections per day (for example, Brinley Buckley et al., 2017).

Although count-based methods for monitoring unmarked organisms are attractive because of their relatively low cost per observation, and ready scalability to high resolution data collection, there are analytical challenges. Population models traditionally require explicit estimation of the rates of birth, death, immigration, and emigration in a population. When individuals are unmarked, and their fates unknown, as is often the case with emerging methods such as camera trapping, births may be confounded with immigration, and deaths with emigration. Given population counts, and assumptions about the relationship between current abundance and rates of immigration and emigration, population dynamics parameters can be estimated from the temporal evolution of count-based population abundance estimates. However, such approaches are not necessarily practical for large count datasets with fine temporal resolution as current statistical methodology for Bayesian analysis of so-called open populations (that is, populations are not demographically closed and new individuals may be added, and individuals may be removed) requires the explicit simulation of the latent immigration and emigration variables (Zipkin et al., 2017). As the spatial and temporal resolutions of available data increase, simulation of the latent states becomes prohibitively expensive to compute and store. Herein we address the problem of scalability by developing a method to estimate high-resolution population change.

As defined in Section 2, *Convolution Likelihood Ecological Abundance Regression (CLEAR)* is an integrated open population model built from the ground up with scalability in mind. Our construction of the model begins with a differential equation for the mean population parameters that defines a mechanistic population dynamics model. We then define suitable likelihoods for the related stochastic difference equation, which defines statistical regression models for high-resolution point count data, low resolution census data, and interview survey data. Improvements in scalability are enabled by using the convolution of the binomial and Poisson distributions as the likelihoods for the count data, which allows marginalization of the unobserved emigration and immigration states. We establish that both the marginal and conditional distributions of this model are Poisson binomial convolutions, so the same likelihood family can be used for initial observations (not conditional on a previous observation) and subsequent observations. The resulting model has similarities with standard integer-valued auto-regressive (INAR) time series (Mckenzie, 1988; Steutel et al., 1983); however, our primary interest lies in a regression analysis of the population parameters and less in traditional time

series analysis. Furthermore, our proposed approach allows for inhomogenous temporal resolutions—unlike standard INAR time series—and can integrate multiple data sources to reduce estimation uncertainty of immigration and emigration. CLEAR also enables posterior predictions of unobserved quantities of interest that can be derived from the immigration and emigration functions, such as predicting aggregate immigration during a given time period (for example, migration).

Herein, we demonstrate how estimating population abundance based on high-frequency count data provides unique insight into the temporal and spatial dynamics of a population that has traditionally proven difficult to study. In the United States, hunting is a popular outdoor recreational activity that creates conservation, social, and economic benefits. Wildlife agencies create public hunting opportunity by buying or leasing land and opening it to public hunting (Nebraska Game & Parks Commission, 2018-05-09a). Land acquisition and rental is expensive and agency budgets are finite. There is consequently a need among public lands managers for tools to assess which prospective lands provide the greatest return in hunter participation. Public land managers have long assessed hunter participation through mail surveys, and more recently through online surveys that ask respondents about what lands they used, what if any game they harvested, and how satisfied they were with their experience (U.S. Fish and Wildlife Service, 2016). Such efforts have vastly increased our understanding of hunter perceptions and motivations. If we are to make informed predictions of a potential public property’s likely utility, however, we must consider hunter habitat decisions in the same detail that we consider any other species with specific habitat needs moving through a landscape (Stedman et al., 2004). The Nebraska Hunter Survey (Section 3) was designed to assess when and where hunters distributed themselves across a diverse public lands-cape. The Nebraska Hunter Survey is the largest in-person survey of recreational hunting known to the authors and treated hunters as a mobile, unmarked population whose abundance varied in space and time, much like a migratory bird or emergent insect population.

Section 4 presents CLEAR analyses of hunter turnout at almost 100 public hunting sites in Southwest Nebraska; Section 4.3 quantifies site-specific use on a daily level (that is, relative abundance estimates), while Section 4.4 estimates the size of the hunter population in the Southwest region by day aggregated over all surveyed public hunting sites and corrected for multiple counting of hunters making trips to multiple sites in a single day (that is, a daily population census). Our substantive analysis has the potential to inform and affect decision making in the management of public land across the country. Section 5 provides concluding remarks.

## 2 Convolution likelihood ecological abundance regression

In this section we develop a regression model based on a convolution likelihood that is inspired by ecological population processes—immigration and emigration. Our methodological developments were motivated and necessitated by the research questions related to the Nebraska Hunter Survey data (Section 3.1), but other potential applications of

this new model class span beyond modeling population abundance and include the modeling of any integer-valued longitudinal panel data.

In the ensuing sections we assume a panel for sites  $s$  on days  $d$  with intraday observations at times  $t$ . The foci of our proposed method are estimating site and day-specific base immigration and emigration rates using geospatial and temporal covariates such as land cover, ownership, accessibility, month, day of week and other seasonalities, and in estimating the intraday patterns of immigration and emigration.

Once the immigration and emigration rates are estimated, we can derive related, but potentially unobserved, quantities of interest such as the cumulative immigration during an arbitrary time period.

## 2.1 Mechanistic model

### General model

We define a model for the expected abundance of an open population with time-varying immigration rate  $i_{s,d}(t) \geq 0$  and emigration rate  $e_{s,d}(t) \geq 0$ , where the subscripted  $s$  and  $d$  indicate a survey site and day, respectively. Our model specification assumes that immigration does not depend on the current abundance of the existing population, but that emigration does. An intuitive justification for our model design is that immigration can occur even when there are currently no individuals in the population, but no emigration can occur while there are no individuals in the population. Denoting the expected abundance of the population at time of day  $t$  by  $\lambda_{s,d}(t)$ , the differential equation describing the population dynamics is

$$\begin{aligned}\frac{\partial}{\partial t}\lambda_{s,d}(t) &= i_{s,d}(t) - \lambda_{s,d}(t) \cdot e_{s,d}(t), \\ \lambda_{s,d}(t_0) &= \lambda_{s,d,0}.\end{aligned}\tag{1}$$

Given that (1) is a first-order linear differential equation, its solution can be obtained by integration,

$$\lambda_{s,d}(t) = \lambda_{s,d}(t_0) \cdot e^{-\int_{t_0}^t e_{s,d}(v)dv} + \int_{t_0}^t i_{s,d}(u) \cdot e^{-\int_u^t e_{s,d}(v)dv} du.\tag{2}$$

### Analytic solution of the differential equation

With the weak restrictions that the immigration function  $i_{s,d}(t)$  be piecewise constant, or linear with respect to time  $t$ , and that the emigration function  $e_{s,d}(t)$  be piecewise constant with respect to time  $t$ , there exists a simple analytic solution of the integrals in (2), which is a crucial requirement for computationally efficient statistical parameter estimation.

Suppose we partition the modeled time period  $[t_0, T)$  into  $n$  intervals  $I_1 := [t_0, t_1)$ ,  $I_2 := [t_1, t_2)$ ,  $\dots$ ,  $I_n := [t_{n-1}, T)$ , and that the immigration and emigration functions are constant or linear (immigration only) within each such interval. The immigration

function  $i_{s,d}(t)$  can then be written as

$$i_{s,d}(t) = i_{s,d,0} \left( \sum_{i=1:n} \mathbf{1}_{I_i}(t)(a_i + b_i(t - t_{i-1})) \right), \tag{3}$$

while the emigration function  $e_{s,d}(t)$  can be written as

$$e_{s,d}(t) = e_{s,d,0} \left( \sum_{i=1:n} \mathbf{1}_{I_i}(t)c_i \right), \tag{4}$$

where the several  $a_i$  and  $c_i$  (for  $i = 0 : n$ ) denote the normalized intercepts and the several  $b_i$  (for  $i = 1 : n$ ) denote the normalized slopes of the functions in interval  $I_i$ . Note that the assumption that the immigration function  $i_{s,d}(t)$  be piecewise continuous requires that  $b_i = \frac{a_{i+1} - a_i}{t_{i+1} - t_i}$ .

For time  $t$  in interval  $I_i$ , and using the piecewise linear and constant variants of the immigration and emigration functions, the solution of the integral in (2) is

$$\begin{aligned} \lambda_{s,d}(t) &= \lambda_{s,d,0} \cdot e^{-e_{s,d,0}(\sum_{j<i} c_j(t_j - t_{j-1}) + c_i(t - t_{i-1}))} \\ &\quad + \sum_{j \leq i} \int_{t_{j-1}}^{\min(t,t_j)} i_{s,d,0} \cdot (a_j + b_j(u - t_{j-1})) \\ &\quad \cdot e^{-e_{s,d,0}(c_j(\min(t,t_j) - u) + \sum_{k=(j+1):i} c_k(\min(t,t_k) - t_{k-1}))} du \\ &= \lambda_{s,d,0} \cdot e^{-e_{s,d,0}(\sum_{j<i} c_j(t_j - t_{j-1}) + c_i(t - t_{i-1}))} \\ &\quad + \sum_{j \leq i} i_{s,d,0} \cdot e^{-e_{s,d,0}(\sum_{k=(j+1):i} c_k(\min(t,t_k) - t_{k-1}))} \\ &\quad \cdot \int_{t_{j-1}}^{\min(t,t_j)} (a_j + b_j(u - t_{j-1})) \cdot e^{-e_{s,d,0}c_j(\min(t,t_j) - u)} du \\ &= \lambda_{s,d,0} \cdot e^{-e_{s,d,0}(\sum_{j<i} c_j(t_j - t_{j-1}) + c_i(t - t_{i-1}))} \\ &\quad + \sum_{j \leq i} i_{s,d,0} \cdot e^{-e_{s,d,0}(\sum_{k=(j+1):i} c_k(\min(t,t_k) - t_{k-1}))} \\ &\quad \cdot \left( \frac{e^{-e_{s,d,0}c_j(\min(t,t_j) - u)} \cdot ((c_j e_{s,d,0})(a_j + b_j(u - t_{j-1})) - b_j)}{c_j^2 e_{s,d,0}^2} \right) \Bigg|_{u=t_{j-1}}^{u=\min(t,t_j)}. \end{aligned} \tag{5}$$

While (5) is rather lengthy, the analytic solution of the integral eliminates the need for numeric integration, and as a result, the differential equation in (1) can be solved using only basic mathematical operations.

## 2.2 Binomial Poisson convolution likelihoods

The binomial distribution is a natural choice to model abundance of a surviving population, while immigration can modeled by a Poisson distribution (Zipkin et al., 2014, 2017; Abadi et al., 2010; Schaub and Abadi, 2011). We derive several key characteristics

of the convolution of a binomial and Poisson distribution from which we propose a new ecological regression model.

Throughout the text, we use the notation

$$N \sim BPC(n, p, \lambda) \tag{6}$$

to denote a random variable  $N = X + Y$  that follows the distribution of the sum of two independent random variables  $X \sim Bin(n, p)$  and  $Y \sim Pois(\lambda)$  with parameters  $n \in \mathbb{N}_0$ ,  $p \in [0, 1]$ , and  $\lambda \geq 0$ .

The mean and variance of a random variable  $N \sim BPC(n, p, \lambda)$  are

$$\mathbb{E}(N) = np + \lambda, \text{ and} \tag{7}$$

$$\mathbb{V}(N) = np(1 - p) + \lambda. \tag{8}$$

Trivial special cases of the BPC distribution include the Dirac distribution  $\delta_n$  for any  $n \in \mathbb{N}_0$ :  $\delta_n = BPC(n, 1, 0)$ ; the Poisson distribution  $Pois(\lambda)$  for any  $\lambda \geq 0$ :  $Pois(\lambda) = BPC(0, p, \lambda)$  for any choice of  $p \in [0, 1]$ ; and the binomial distribution  $Bin(n, p)$  for any  $n \in \mathbb{N}_0$  and  $p \in [0, 1]$ :  $Bin(n, p) = BPC(n, p, 0)$ .

**Closure under BPC evolution**

Suppose that  $X$  and  $Y$  are independent non-negative integer-valued random variables with probability mass functions  $p_X$  and  $p_Y$ , respectively. With  $S = X + Y$ ,  $U|X = Bin(X, p)$ ,  $V|Y = Bin(Y, p)$ ,  $W|S = Bin(S, p)$ , we show that  $U + V$  and  $W$  have the same distribution, so applying binomial survival to the sum of two independent random variables leads to the same outcome as applying the survival to each random variable separately and then adding them together.

$$\begin{aligned} p_{U+V}(k) &= \sum_{i=0:k} p_U(i)p_V(k-i) \\ &= \sum_{i=0:k} \left( \sum_x p_X(x) \binom{x}{i} p^i (1-p)^{x-i} \right) \left( \sum_y p_Y(y) \binom{y}{k-i} p^{k-i} (1-p)^{y-(k-i)} \right) \\ &= \sum_{i=0:k} \sum_x \sum_y p_X(x)p_Y(y)p^k (1-p)^{(x+y)-k} \binom{x}{i} \binom{y}{k-i} \\ &= \sum_x \sum_y p_X(x)p_Y(y)p^k (1-p)^{(x+y)-k} \left( \sum_{i=0:k} \binom{x}{i} \binom{y}{k-i} \right) \\ &= \sum_x \sum_y p_X(x)p_Y(y)p^k (1-p)^{(x+y)-k} \binom{x+y}{k} \\ &= \sum_s p_S(s) \binom{s}{k} p^k (1-p)^{s-k} \\ &= p_W(k) \end{aligned} \tag{9}$$

In this derivation we utilize Vandermonde's identity in the key transformation  $\sum_{i=0:k} \binom{x}{i} \binom{y}{k-i} = \binom{x+y}{k}$ .

In particular, this general result applies to the BPC distribution, given that it is the sum of independent binomially-distributed and Poisson-distributed random variables; that is,

$$K \sim BPC(n, pq, q\lambda), \quad (10)$$

if  $N \sim BPC(n, p, \lambda)$  and  $K|N \sim Bin(N, q)$ , assuming as well-known that if  $A \sim Pois(\lambda)$  and  $B|A \sim Bin(A, q)$ , then  $B \sim Pois(q\lambda)$ , and if  $A \sim Bin(n, p)$  and  $B|A \sim Bin(A, q)$ , then  $B \sim Bin(n, pq)$ .

Furthermore, the BPC distribution is closed under convolution with a Poisson distribution given that the BPC distribution is constructed as the convolution of a binomial and Poisson distribution, and the sum of two independent Poisson-distributed random variables is also Poisson-distributed. As a result,

$$M \sim BPC(n, pq, q\lambda + \gamma), \quad (11)$$

if  $N \sim BPC(n, p, \lambda)$  and  $M|N \sim BPC(N, q, \gamma)$ .

### Connections to the INAR literature

The literature on INAR models and self-decomposability of discrete random variables (for example, Mckenzie, 1988; Steutel and van Harn, 1979; Steutel et al., 1983) uses different nomenclature and notation:

- Binomial “survival” (Section 2.2) is called binomial “thinning” in the INAR literature;
- The INAR notation for the evolution  $M|N \sim Bin(N, q) + Pois(\lambda)$  is  $M = q \circ N + I$ , where  $I \sim Pois(\lambda)$ .

Here it is relevant to present both notations to link the mostly separate bodies of literature on probability theory and stochastic processes, with that of quantitative ecological analysis.

### 2.3 Convolution likelihood ecological abundance regression

This section combines results from the previous sections and formally defines a statistical regression model for longitudinal panel data for counts (non-negative integers). In a first step, we establish that the proposed model is indifferent to the temporal resolution of the data; in particular, data with inhomogeneous temporal resolution data can be easily incorporated in CLEAR analysis.

**Robustness to inhomogeneous temporal resolution**

Let the time evolutions of the stochastic abundance process from time  $t_i$  to any time  $t_j > t_i$  be defined as follows.

$$N_{s,d,t_j} | N_{s,d,t_i} \sim BPC \left( N_{s,d,t_i}, e^{-\int_{t_i}^{t_j} e_{s,d}(v) dv}, \int_{t_i}^{t_j} i_{s,d}(u) \cdot e^{-\int_u^{t_j} e_{s,d}(v) dv} du \right). \quad (12)$$

We show that the evolutions in (12) are consistent under iterative application; first evolving  $N_{s,d,t}$  from time  $t_i$  to  $t_{i+1}$ , and then from  $t_{i+1}$  to  $t_{i+2}$  leads to the same result for  $N_{s,d,t_{i+2}}$  as evolving  $t_i$  directly to  $t_{i+2}$ . Let

$$N_{s,d,t_{i+1}} | N_{s,d,t_i} \sim BPC \left( N_{s,d,t_i}, e^{-\int_{t_i}^{t_{i+1}} e_{s,d}(v) dv}, \int_{t_i}^{t_{i+1}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+1}} e_{s,d}(v) dv} du \right) \quad (13)$$

and

$$N_{s,d,t_{i+2}} | N_{s,d,t_{i+1}} \sim BPC \left( N_{s,d,t_{i+1}}, e^{-\int_{t_{i+1}}^{t_{i+2}} e_{s,d}(v) dv}, \int_{t_{i+1}}^{t_{i+2}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+2}} e_{s,d}(v) dv} du \right). \quad (14)$$

Applying the result on BPC evolution in (11) to these two evolution steps we obtain the desired result,

$$\begin{aligned} N_{s,d,t_{i+2}} | (N_{s,d,t_{i+1}} | N_{s,d,t_i}) &\sim BPC \left( N_{s,d,t_i}, e^{-\int_{t_i}^{t_{i+1}} e_{s,d}(v) dv} e^{-\int_{t_{i+1}}^{t_{i+2}} e_{s,d}(v) dv}, \right. \\ &\quad \left( \int_{t_i}^{t_{i+1}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+1}} e_{s,d}(v) dv} du \right) \cdot e^{-\int_{t_{i+1}}^{t_{i+2}} e_{s,d}(v) dv} \\ &\quad \left. + \int_{t_{i+1}}^{t_{i+2}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+2}} e_{s,d}(v) dv} du \right) \\ &= BPC \left( N_{s,d,t_i}, e^{-\int_{t_i}^{t_{i+1}} e_{s,d}(v) dv - \int_{t_{i+1}}^{t_{i+2}} e_{s,d}(v) dv}, \right. \\ &\quad \left. \int_{t_i}^{t_{i+1}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+1}} e_{s,d}(v) dv - \int_{t_{i+1}}^{t_{i+2}} e_{s,d}(v) dv} du \right. \\ &\quad \left. + \int_{t_{i+1}}^{t_{i+2}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+2}} e_{s,d}(v) dv} du \right) \\ &= BPC \left( N_{s,d,t_i}, e^{-\int_{t_i}^{t_{i+2}} e_{s,d}(v) dv}, \right. \\ &\quad \left. \int_{t_i}^{t_{i+2}} i_{s,d}(u) \cdot e^{-\int_u^{t_{i+2}} e_{s,d}(v) dv} du \right) \\ &\sim N_{s,d,t_{i+2}} | N_{s,d,t_i}. \end{aligned} \quad (15)$$

As a result, if  $N_{s,d,t_i}$  is known or otherwise BPC-distributed, any future  $N_{s,d,t_j}$  ( $t_j > t_i$ ) is also marginally BPC-distributed. Consider the relevant special case that  $N_{s,d,t_0}$  is unobserved and modeled as a latent Poisson-distributed random variable with mean  $\lambda_{s,d,0}$ ,

$$N_{s,d,t_0} \sim Pois(\lambda_{s,d,0}) = BPC(0, 0, \lambda_{s,d,0}). \tag{16}$$

BPC evolution to time  $t$  yields

$$N_{s,d,t} | N_{s,d,t_0} \sim BPC \left( N_{s,d,t_0}, \mathbf{e}^{-\int_{t_0}^t e_{s,d}(v)dv}, \int_{t_0}^t i_{s,d}(u) \cdot \mathbf{e}^{-\int_u^t e_{s,d}(v)dv} du \right),$$

and the marginal distribution of  $N_{s,d,t}$  follows as

$$\begin{aligned} N_{s,d,t} &\sim BPC \left( 0, 0, \lambda_{s,d,0} \cdot \mathbf{e}^{-\int_{t_0}^t e_{s,d}(v)dv} + \int_{t_0}^t i_{s,d}(u) \cdot \mathbf{e}^{-\int_u^t e_{s,d}(v)dv} du \right) \\ &= BPC(0, 0, \lambda_{s,d}(t)) \\ &= Pois(\lambda_{s,d}(t)). \end{aligned} \tag{17}$$

**Model for intraday data**

High-frequency intraday data feeds into the analysis conditional on the preceding observation to maximize information gain. To make Bayesian analysis feasible, we reduce the computational burden of likelihood evaluations by eliminating the need for numerical evaluation of the integrals in (12). Our model definition allows for inhomogeneous temporal resolution, which is typical in ecological applications (see, for example, Stuber et al., 2017).

We denote the intraday observation times by  $t_0 < t_1 < t_2 < \dots < t_n$ . Suppose that the immigration function  $i_{s,d}(t)$  is linear in each observation period  $I_i := [t_i, t_{i+1})$  and that the emigration function  $e_{s,d}(t)$  is constant in each observation period  $I_i$  as in Section 2.1; recall the notation  $i_{s,d}(t) = i_{s,d,0} \cdot (a_i + b_i t)$  and  $e_{s,d}(t) = e_{s,d,0} \cdot c_i$  for  $t \in I_i$ . Then the results from Section 2.1 apply, and the conditional likelihoods for  $N_{s,d,t_{i+1}} | N_{s,d,t_i}$  simplify—in terms of computational burden—to

$$\begin{aligned} N_{s,d,t_{i+1}} | N_{s,d,t_i} &\sim BPC \left( N_{s,d,t_i}, \right. \\ &\quad \mathbf{e}^{-e_{s,d,0} \cdot c_i \cdot (t_{i+1} - t_i)}, \\ &\quad \left. \frac{i_{s,d,0}}{e_{s,d,0}^2 \cdot c_i^2} \cdot \left( (a_i c_i e_{s,d,0} + b_i c_i e_{s,d,0} (t_{i+1} - t_i) - b_i) \right. \right. \\ &\quad \left. \left. - \mathbf{e}^{-e_{s,d,0} c_i (t_{i+1} - t_i)} \cdot (a_i c_i e_{s,d,0} - b_i) \right) \right). \end{aligned} \tag{18}$$

This model is very similar to an INAR(1) model with binomial thinning and Poisson residuals; however we do not assume homogeneous temporal resolution, and the

relationships between covariates  $\mathbf{x}_{s,d}$  and immigration and emigration are of interest (regression coefficients  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ ) instead of just the mean survival probabilities. These regression effects are included as

$$i_{s,d,0} = \exp(\mathbf{x}'_{s,d}\boldsymbol{\alpha}), \text{ and} \tag{19}$$

$$e_{s,d,0} = \exp(\mathbf{x}'_{s,d}\boldsymbol{\beta}), \tag{20}$$

where the log link function guarantees positivity. Appropriate constraints for the several  $a_i$ ,  $b_i$  and  $c_i$  ensure likelihood-identifiability of the parameters (more details are in Section 4.1).

The unknown model parameters of this model are  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and the several  $a_i$  and  $c_i$ ; recall that the several  $b_i$  are determined fully by the continuity assumption for  $i_{s,d}(t)$ , which requires that  $b_i = \frac{a_{i+1}-a_i}{t_{i+1}-t_i}$ .

**Model for low-frequency data**

For data that is collected at most once per day, there is no preceding same-day observation. Such data are modeled conditional on a latent initial population at time  $t_0$ , modeled by a Poisson distribution, where the covariates  $\mathbf{x}_{s,d}$  are included via a log link as

$$\lambda_{s,d,0} = \exp(\mathbf{x}'_{s,d}\boldsymbol{\alpha}_0), \tag{21}$$

$$N_{s,d,t_0} \sim Pois(\lambda_{s,d,0}). \tag{22}$$

It follows from (17) that the observed data at time  $t$  are marginally Poisson-distributed with mean  $\lambda_{s,d}(t)$ . With piecewise linear/constant immigration and emigration in (5), the closed-form likelihood for  $N_{s,d,t}$  follows as

$$\begin{aligned} N_{s,d,t} \sim &Pois\left(\lambda_{s,d,0} \cdot \mathbf{e}^{-e_{s,d,0}(\sum_{j<i} c_j(t_j-t_{j-1})+c_i(t-t_{i-1}))}\right. \\ &+ \sum_{j \leq i} i_{s,d,0} \cdot \mathbf{e}^{-e_{s,d,0}(\sum_{k=(j+1):i} c_k(\min(t,t_k)-t_{k-1}))} \\ &\left. \cdot \left(\frac{\mathbf{e}^{-e_{s,d,0}c_j(\min(t,t_j)-u)} \cdot ((c_j e_{s,d,0})(a_j + b_j(u - t_{j-1})) - b_j)}{c_j^2 e_{s,d,0}^2}\right)\right) \Bigg|_{u=t_{j-1}}^{u=\min(t,t_j)}. \end{aligned} \tag{23}$$

Availability of a closed-form solution for the likelihood of  $N_{s,d,t}$  is key for efficient estimation as it eliminates the need for computationally-intensive modeling of the latent time evolutions.

The unknown parameters for this low-frequency model are  $\boldsymbol{\alpha}_0$  in addition to those of the high-frequency model ( $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and the several  $a_i$  and  $c_i$ ).

**Model for interview data**

Exit interview data supplies individuals’ hunt duration, which is a transformed realization of emigration. Indeed, moment-matching the mechanistic model (1) implies a probability distribution for hunt duration  $D_{s,d,t}$ : the probability that a hunting trip to site  $s$  on date  $d$  that started at time  $t$  lasts at least duration  $\Delta$  (some unit of time) equals the expected abundance at time  $t + \Delta$  conditional on abundance at hunt start being one and no immigration during the trip.

$$\begin{aligned}
 P(D_{s,d,t} \geq \Delta) &= \mathbb{E}(\mathbf{1}_{D_{s,d,t} \geq \Delta}) \\
 &= \mathbb{E} \left( N_{s,d,t+\Delta} | N_{s,t,d} = 1, \int_t^{t+\Delta} i_{s,d}(u) du = 0 \right) \\
 &= e^{-\int_t^{t+\Delta} e_{s,d}(v) dv} \\
 &= e^{-e_{s,d,0} \left( \sum_{\{j: t_{j-1} \leq t+\Delta, t_j \geq t\}} c_j (\min(t_j, t+\Delta) - \max(t_{j-1}, t)) \right)} \tag{24}
 \end{aligned}$$

The probability density function (p.d.f.) of hunt duration  $D_{s,d,t}$  follows as

$$\begin{aligned}
 p(\Delta) &= \frac{\partial}{\partial \Delta} P(D_{s,d,t} \leq \Delta) = \frac{\partial}{\partial \Delta} (1 - P(D_{s,d,t} \geq \Delta)) \\
 &= \frac{\partial}{\partial \Delta} \left( 1 - e^{-\int_t^{t+\Delta} e_{s,d}(v) dv} \right) \\
 &= -e^{-\int_t^{t+\Delta} e_{s,d}(v) dv} \times (-e_{s,d}(t + \Delta)) \\
 &= e^{-\int_t^{t+\Delta} e_{s,d}(v) dv} \times e_{s,d}(t + \Delta) \\
 &= e^{-e_{s,d,0} \left( \sum_{\{j: t_{j-1} \leq t+\Delta, t_j \geq t\}} c_j (\min(t_j, t+\Delta) - \max(t_{j-1}, t)) \right)} \times e_{s,d,0} \times c_{j'}, \tag{25}
 \end{aligned}$$

with  $j'$  such that  $t_{j'-1} \leq (t + \Delta) < t_{j'}$ .

The resulting distribution is an exponential distribution with inhomogeneous rate  $e_{s,d}(t)$ . The unknown parameters for the hunt duration model are those of the emigration sub-model of the high-frequency model ( $\beta$  and the several  $c_i$ ).

**Remark on Identifiability**

Without the interview/hunt duration component model, we would be estimating two unobserved quantities, immigration and emigration, using a response variable that is the difference of the two, which could lead to identification problems as they appear in analyses of the  $N$ -mixture model (Royle, 2004; Dail and Madsen, 2011; Barker et al., 2018; Kéry, 2018; Link et al., 2018; Knape and Korner-Nievergelt, 2016). However, our interview component model provides a direct estimate of the emigration function, which helps the point count component models identify the correct level of immigration (see the supplementary materials for details on model validation, Gruber et al., 2019).

### 3 The Nebraska hunter survey

North American wildlife managers are charged with conserving wildlife populations and providing recreational opportunities for hunters and other recreational land users. Conservation practitioners may draw from an expansive body of scientific and institutional knowledge when making decisions for wildlife populations. In contrast, a lack of information on how hunters respond to varying hunting access landscapes is a significant challenge to creating effective public lands management strategies (Ryan and Shaw, 2011). Hunters, like the animals they hunt, navigate cues of habitat suitability at multiple scales when choosing how and when to move through a landscape (Mecozzi and Guthery, 2008). Game animals may only be legally hunted during defined seasons intended to balance hunter satisfaction and biological integrity (Nebraska Game & Parks Commission, 2018-05-09b). For example, ducks may be hunted during a portion of their fall migration, but not during the breeding season. Game animals are also distributed across the landscape according to their own habitat needs and decisions, and hunters ostensibly choose to hunt in times and places where they expect to encounter target game animals (Lone et al., 2014). One of the intended purposes of the Nebraska Hunter Survey was to understand how hunter use of different public lands hunting opportunities varied in space and time so that public lands managers could make informed decisions about future land leases and purchases.

#### 3.1 Research questions addressed in this manuscript

In this manuscript, we ask:

1. Section 4.3: What is site-specific daily use measured by the number of hunting trips made to each site? This requires an estimate of the cumulative arrivals of cars (and by extension, hunters) for each day and site.
2. Section 4.4: How many hunters are using public hunting sites on each day? We need to account for hunters making multiple hunting trips on a single day when aggregating the number of site-specific hunting trips as well as for the party size per car.

#### 3.2 Survey methodology

From 2014-2017, the Nebraska Hunter Survey collected data assessing hunter activity at 512 unique sampling locations across the State of Nebraska, USA covering seven focal regions (Figure 1). Sampling locations were areas of land open to public hunting through permanent public ownership (for example, state wildlife management areas), or leased for public hunting through public-private partnerships (Nebraska Game & Parks Commission, 2018-05-09a). The Nebraska Hunter Survey collected three datasets: high-frequency sub-sample counts (time-lapse images), low-frequency full-population counts, and exit interviews.

Sampling locations were delineated by the Public Land Survey System at the Direction-Township-Range-Section resolution (U.S. Geological Survey, 2018). Each sampling

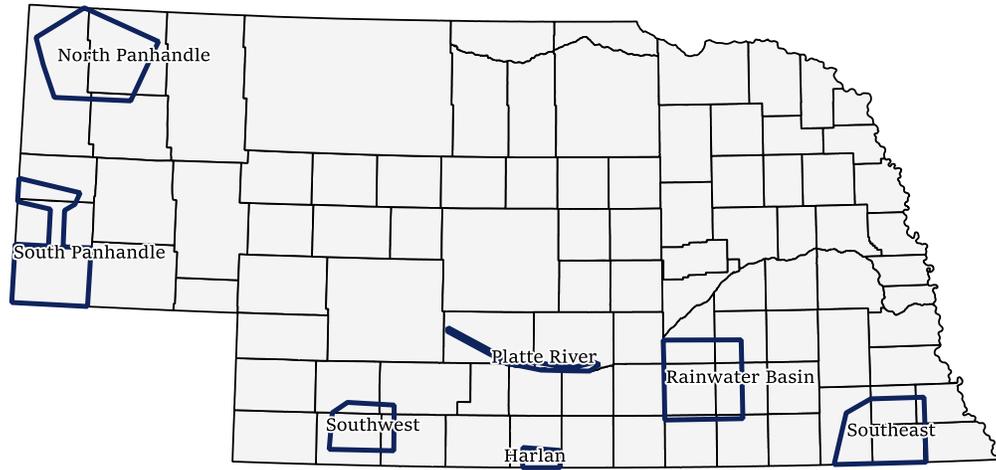


Figure 1: Spatial distribution of seven focal regions in which multiple public accessible lands were surveyed for use by hunters.

location thus had a maximum area of about 259 ha (one “section”), although most were smaller and many were adjacent to other sampling locations. Hunter data was collected from 1 September to 31 January of the following year.

### 3.3 Datasets

#### High-frequency intraday counts

We deployed trail cameras to create a high-frequency record of the number of vehicles observed at public access hunting sites. Trail cameras were attached to elevated structures (for example, trees, poles) on the perimeter of sites at a height of 4-5 meters, and were pointed toward parking areas. Cameras recorded an image every five minutes from sunrise to sunset. Images were processed in *Timelapse* (Greenberg, 2015), and information on image quality (for example, clear, fogged, corrupted), and the number of people and vehicles visible recorded. Counts represent a sample of the full population present at the sampling site (one section), given the limited field of view of a camera. We collected a total of 11,608,989 time-lapse images from 229 sites.

#### Low-frequency full-population counts

Full population counts were conducted every day during the sampling period, except on days when weather rendered roads impassible. Morning sampling routes began at the start of legal hunting hours, 30 minutes before sunrise, and continued until noon. Afternoon routes began between 12:30 and 13:30, according to sunset times, and continued until 30 minutes after sunset. We recorded the total number of vehicles present at each site, and vehicle origin (local Nebraska county, other Nebraska county, out-of-

state, unknown) based on license plates. The count schedule randomized start times (either morning or afternoon), start locations (2-4 per study region), and route direction (clockwise or counter clockwise). Each sampling location was only sampled once on any given day. We collected a total of 110,199 point count records from 562 sites.

### Exit interviews

Technicians interviewed public lands users over the age of 19 opportunistically before, during, and after point counts. The interviewer first obtained consent from the interview subject, then asked them what activity they were pursuing, what species if hunting, and a range of demographic questions (age, sex, home zip code). The interview was reactive, routing interview subjects through questions according to their responses (for example, a duck hunter would be asked what species of ducks they harvested, but a cyclist would not). One interview was recorded per hunting party, the social group of one or more recreationists pursuing an outdoor activity on public lands. We collected a total of 3,088 interviews at 459 sites. Methods involving human subjects were approved by The University of Nebraska-Lincoln Committee on Research Involving Human Subjects (IRB # 20120912892EX).

## 4 Application study: estimating daily site use

Site use is a key performance indicator for evaluating and managing public access hunting sites. We define use of a hunting site  $s$  on day  $d$  as the number of cars arriving at that site during that day (number of hunting trips made to a site during a day), denoted by  $SN_{s,d}$ . A key challenge in this analysis is that  $SN_{s,d}$  is unobserved, as our point counts only provide snapshots of the car populations at discrete sampling times  $t$ , denoted  $N_{s,d}(t)$ , but not cumulative counts of arriving cars throughout each day.

The car populations  $N_{s,d}(t)$  at site  $s$  during day  $d$  at times  $t$  are modeled as open populations—cars can arrive and leave at any time. To relate the observed data,  $N_{s,d}(t)$ , to our variable of interest,  $SN_{s,d}$ , we note that both are two different measurements of the same underlying intraday immigration and emigration processes that describe the flow of cars arriving at and departing from a site throughout the day. We estimate the parameters of the immigration and emigration functions using Convolution Likelihood Ecological Abundance Regression (CLEAR) as introduced in Section 2.3; once the immigration function is estimated, the cumulative car counts  $SN_{s,d}$  can be trivially derived by integration of the immigration function over time.

The benefits of CLEAR over previous methods (for example, Zipkin et al., 2014) are twofold. First, explicit simulation of the latent immigration and emigration states would be prohibitively memory-intensive for our high-resolution data. Second, CLEAR enables estimation of cumulative immigration  $\int_{t_i}^{t_j} i_{s,d}(u)du$  during arbitrary time periods  $[t_i, t_j)$ , while the latent immigration states of previous models represent immigration net of emigration (not gross immigration), and are only available on the temporal resolution explicitly being simulated.

### 4.1 Model specification

#### Likelihood

The high-frequency time-lapse photography data  $C_{s,d,t}$  (see Section 3.3) are modeled as in (18) with the minor modification that the base immigration rate  $i_{s,d,0}$  be scaled by a site-specific factor  $v_s$  to account for the limited viewing angle of the camera,

$$C_{s,d,t_{i+1}} | C_{s,d,t_i} \sim BPC \left( C_{s,d,t_i}, \right. \\ \left. e^{-e_{s,d,0} \cdot c_i \cdot (t_{i+1} - t_i)}, \right. \\ \left. v_s \cdot \frac{i_{s,d,0}}{e_{s,d,0}^2 \cdot c_i^2} \cdot \left( (a_i c_i e_{s,d,0} + b_i c_i e_{s,d,0} (t_{i+1} - t_i) - b_i) \right. \right. \\ \left. \left. - e^{-e_{s,d,0} c_i (t_{i+1} - t_i)} \cdot (a_i c_i e_{s,d,0} - b_i) \right) \right). \quad (26)$$

The low-frequency full-population point counts  $N_{s,d,t}$  (see Section 3.3) are modeled as in (23),

$$N_{s,d,t} \sim Pois \left( \sum_{j \leq i} i_{s,d,0} \cdot e^{-e_{s,d,0} (\sum_{k=(j+1):i} c_k (\min(t, t_k) - t_{k-1}))} \right. \\ \left. \cdot \left( \frac{e^{-e_{s,d,0} c_j (\min(t, t_j) - u)} \cdot ((c_j e_{s,d,0}) (a_j + b_j (u - t_{j-1})) - b_j)}{c_j^2 e_{s,d,0}^2} \right) \Bigg|_{u=t_{j-1}}^{u=\min(t, t_j)} \right).$$

The hunt duration interview survey data  $D_{s,d,t}$  (see Section 3.3) are modeled as in (25), with the appropriate correction for right-censored data collection, given that the sampling protocol specified no data be collected after cutoff time  $T = t_n$  at the end of legal hunting hours,

$$p_{D_{s,d,t}}(\Delta) = \frac{e^{-e_{s,d,0} (\sum_{\{j: t_{j-1} \leq t + \Delta, t_j \geq t\}} c_j (\min(t_j, t + \Delta) - \max(t_{j-1}, t)))} \times e_{s,d,0} \times c_{j'}}{1 - e^{-e_{s,d,0} (\sum_{\{j: t < t_j \leq T\}} c_j (t_j - \max(t_{j-1}, t))})}. \quad (27)$$

In (27),  $j'$  is chosen such that  $t_{j'-1} \leq t + \Delta < t_{j'}$ .

We allow for different log-likelihood weights for the three component models, to counteract potentially unbalanced input data sets.

Log link functions map the linear predictors with regression coefficients  $\alpha$  and  $\beta$  to the site and day-specific base immigration and emigration rates  $i_{s,d,0}$  and  $e_{s,d,0}$ ,

$$i_{s,d,0} = \exp(\mathbf{x}'_{s,d} \alpha), \text{ and} \quad (28)$$

$$e_{s,d,0} = \exp(\mathbf{x}'_{s,d} \beta); \quad (29)$$

a logit link maps the linear predictor with regression coefficients  $\gamma$  to the site-specific camera coverage rate  $v_s$ ,

$$v_s = \text{logit}^{-1}(\mathbf{c}'_s \gamma). \tag{30}$$

The corresponding covariate vectors are denoted by  $\mathbf{x}_{s,d}$  and  $\mathbf{c}_s$ . Initial populations  $\lambda_{s,d,0}$  as in (23) are excluded from our model, given that no camping sites or other locations with expected overnight populations were surveyed.

The main purpose of the high-frequency component model is to improve the estimates of intraday immigration and emigration dynamics, which also improves the predictive power of combined model. In our study, there is little practical utility in predicting from the high-frequency model directly other than to validate the model and method; however, commercial applications such as those focused on real-time inventory management and pricing could benefit substantially from such short-term forecasts.

**Priors**

As a canonical choice, we use standard normal priors for all of the regression coefficients,

$$\gamma \sim N(0, 1), \tag{31}$$

$$\alpha \sim N(0, 1), \text{ and} \tag{32}$$

$$\beta \sim N(0, 1). \tag{33}$$

To confirm that the prior choice has only limited impact on the estimates of site utilization, we also run the model with prior standard variances 0.1 and 10 and report the site use and population estimates from all three models.

The regression formulas for the base levels for immigration and emigration,  $i_{s,d,0}$  and  $e_{s,d,0}$ , respectively, typically contain an intercept term. As a result, if the intraday time levels and intercepts of immigration and emigration (these are the several  $a_i$  and  $c_i$ ) were unbounded, the model would not be likelihood-identifiable: any change in the intercept term could be neutralized by changes in the intraday time levels. We address this by including a “sum to a constant” constraint in the prior for the intraday time levels of immigration and emigration, the several  $a_i$  and  $c_i$ ; no priors are required for the slopes  $b_i = \frac{a_{i+1}-a_i}{t_{i+1}-t_i}$  as these are fully specified conditional on the several  $a_i$ . Given  $n$  piecewise linear/constant time periods  $I_i = [t_{i-1}, t_i]$ ,  $i = 1 : n$ , we the priors for the several  $a_i$  and  $c_i$  are scaled Dirichlet distributions of the form

$$\frac{(a_0, \dots, a_n)}{n + 1} \sim \text{Dirichlet} \left( \frac{\kappa}{n + 1} \times (1, \dots, 1) \right), \text{ and} \tag{34}$$

$$\frac{(c_0, \dots, c_n)}{n + 1} \sim \text{Dirichlet} \left( \frac{\kappa}{n + 1} \times (1, \dots, 1) \right), \tag{35}$$

where  $\kappa > 0$  is a prior parameter (in the analysis below we use  $\kappa = 10$ ). The scaled Dirichlet priors accomplish identifiability of the regression intercepts by normalizing the mean of the intraday time levels to one,  $\frac{1}{n+1} \sum_i a_i = 1$  and  $\frac{1}{n+1} \sum_i c_i = 1$ . The overall immigration and emigration levels are more appropriately modeled by their intercept regression coefficients.

Geospatial covariates	Temporal covariates
<ul style="list-style-type: none"> <li>• Center latitude of section</li> <li>• Center longitude of section</li> <li>• Area (ha) of section</li> <li>• Public area (ha) in section</li> <li>• Public area (ha) in section and surrounding sections</li> <li>• CRP area (ha) in section and surrounding sections</li> <li>• Total road length (m) in section and surrounding sections</li> <li>• Primary entity type of public lands in section</li> <li>• Number of public hunting sites on hunter atlas page</li> </ul>	<ul style="list-style-type: none"> <li>• Study year</li> <li>• Day of week</li> <li>• Month</li> <li>• Federal holiday indicator</li> <li>• Deer hunting season indicators</li> <li>• Pheasant and quail hunting season indicators</li> <li>• Grouse hunting season indicators</li> <li>• Turkey hunting season indicators</li> <li>• Teal hunting season indicators</li> <li>• Duck and coot hunting season indicators</li> </ul>
Geospatial and temporal predictors	
<ul style="list-style-type: none"> <li>• Precipitation</li> <li>• Snowfall</li> </ul>	

Table 1: Geospatial and temporal covariates used to model the base immigration and emigration rates.

## 4.2 Data and implementation

We analyze site use in the Southwest hunting region, which is a popular destination for pheasant and deer hunters in Nebraska. Table 1 lists the geospatial and temporal covariates used to model initial abundance as well as the base immigration and emigration rates,  $i_{s,d,0}$  and  $e_{s,d,0}$ , respectively.

The Southwest Nebraska data contains 21,339 observations of the low-frequency point counts, 915,077 observations of the high-frequency point counts, and 506 observations of hunt duration. We randomly partitioned the data into training and validation data, using one third of the low-frequency data (7,183 observations), 10% of the high-frequency data (91,004 observations), and two thirds of the interview data (321 observations) for training, and the rest for validation. Furthermore, the interview likelihoods were up-weighted by a factor of 5, the low-frequency point count likelihoods received a unit likelihood weight, and the high-frequency point count likelihoods were down-weighted by a factor of 0.1 to counteract the unbalanced data sets.

Analyses were performed in Stan (Carpenter et al., 2017), which allowed the custom definition of our BPC likelihood function. Analyses are based on a posterior sample of 1,000 iterations after a warm-up phase of 500 iterations. Readers familiar with MCMC may note the relatively low number of iterations; this is due to Stan’s use of Hamiltonian MC, which is less prone to convergence problems.

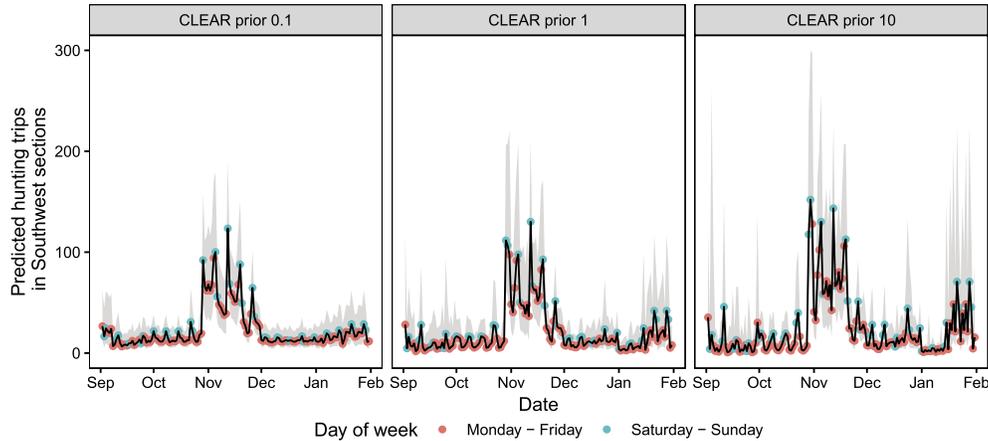


Figure 2: Estimates hunting trips per day to public access sites in all 89 sections; the shaded area shows the 95% credible interval; the black line shows the median estimates.

Model fit of the component models is evaluated in the supplementary materials (Gruber et al., 2019).

### 4.3 Estimated number of hunting trips

We estimate arrivals ( $SN_{s,d}$ ) for 89 sections with public hunting sites in the Southwest region for which we have geospatial covariates and all 152 days  $d$  in the 2016 fall hunting season (September 2, 2016 through January 31, 2017). The estimates shown in Figure 2 are based on posterior samples of the daily arrivals per section.

Of the 89 public hunting sections, 50 are primarily or entirely privately owned and offer 3,418 ha of public access lands, while the other 39 are primarily or entirely state owned and offer 5,145 ha of public access lands. Privately-owned land enrolled in the State’s public access program tended to be enrolled in the Conservation Reserve Program (CRP), which provides pheasant and quail habitat that may attract pheasant and quail hunters (Anderson and David, 1998; Hiller et al., 2015). In contrast, state-owned land in the Southwest study region are mainly riparian forest and river bottoms better suited to turkey, deer, and duck hunting. The estimates shown in Figure 3 confirm this hypothesis, with privately owned public access sites realizing peak turnout during the pheasant and quail hunting season, while state owned sites show comparatively more activity outside of the pheasant and quail hunting season, and highest turnout during the deer hunting season. Over the whole fall 2016 season, the 50 privately-owned sections are estimated to have attracted 1,035 hunting trips, while the 39 state-owned sections are estimated to have attracted 2,161 hunting trips (figures are the sums of the daily median estimates from the CLEAR model with prior variance 1).

Site use is estimated to be substantially higher on weekends than during the week, with weekends estimated to account for about 57% of all hunting trips. Season opening

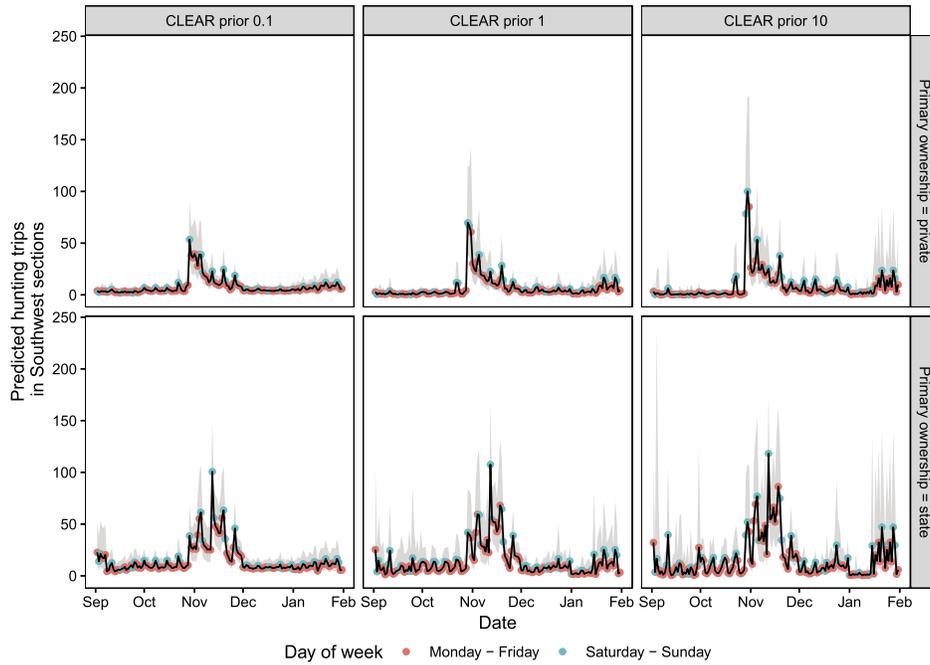


Figure 3: Estimated hunting trips per day to public access sites in all 89 sections by primary ownership type; the shaded area shows the 95% credible interval; the black line shows the median estimates.

weekends carry special significance; the pheasant opening weekend is estimated to have captured 13% of all trips made to privately-owned sites during the entire fall season, and the deer opening weekend is estimated to have captured 7.5% of all trips made to publicly-owned sites. Hunting activity is elevated for about four weeks, starting with the pheasant season opener (estimated hunting trips to the surveyed 89 public access hunting sections in Southwest Nebraska for the four peak Saturdays are 108 hunting trips per day). Outside of the main hunting seasons, estimated use is substantially lower at about 12 hunting trips per day across all sites. In relative numbers, sites are expected to be 9 times as busy on peak Saturdays than on the average off-season day. Additional estimates are in Table 2.

The estimates from the CLEAR models with the three different prior choices generally agree and show similar patterns. As expected, peaks are more pronounced in the model with weaker priors.

#### 4.4 Estimated hunter population

We extend the car population model from the previous section with a normalizing process to only count cars that were not counted at different sites earlier on the same

	Prior 0.1	Prior 1	Prior 10
Trips	3,546	3,197	3,752
Trips to private sites	1,222	1,035	1,250
Trips to state sites	2,324	2,161	2,502
Trips on weekends	1,352	1,384	1,674
Trips Monday-Friday	2,194	1,813	2,078
Trips to private sites on pheasant opening weekend	92	136	178
Trips to state sites on deer opening weekend	157	162	173
Trips during peak season (incl. weekends)	1,629	1,681	2,008
Trips on 4 peak Saturdays	404	433	504
Trips on 4 peak Sundays	239	270	328

Table 2: Estimated number of hunting trips (posterior medians) during the fall 2016 season across all 89 surveyed sites in Southwest Nebraska.

day, and to account for the expected size of the hunting party. Hunters frequently hunt at multiple sites on a single day, which leads to multiple estimated arrivals across all sites, which is relevant and desirable when estimating individual site use, but undesirable when creating a population abundance estimate. The exit interview data (see Section 3.3) contains information on whether a recreationist had visited another site on the same day before arriving at the site of the interview, and also contains information about the size of the surveyed parties, which enables statistical estimation of these quantities.

We estimate two additional models, one for the probability of a hunting trip being the first of the day, and one for the expected size of the hunting party. We denote the response variable for the first trip of the day question of an interview conducted at site  $s$  on day  $d$  and for a hunt start time  $t$  by  $F_{s,d}(t)$  with the convention that  $F_{s,d}(t)$  takes value 1 for “no” and value 0 for “yes.” We estimate a logistic regression model to estimate  $F_{s,d}(t)$  using a subset of the geospatial and temporal covariates listed in Table 1, and modeling the intraday patterns in a piecewise linear fashion as for the immigration function (Section 2.1). Furthermore, we estimate a Poisson regression model to estimate the expected party size  $g_{s,d}(t)$ , again using a subset of the geospatial and temporal covariates listed in Table 1, and modeling the intraday patterns in a piecewise linear fashion as for the immigration function (Section 2.1).

The estimate of the daily total hunter population across all sites  $s$  on day  $d$ , denoted  $TN_d$ , is obtained through integration,

$$TN_d = \sum_s \int_{t_0}^{t_n} i_{s,d}(u) \cdot g_{s,d}(u) \cdot \mathbb{P}(F_{s,d}(u) = 1) du, \quad (36)$$

where  $F_{s,d}(t)$  is an indicator variable for a trip to site  $s$  on day  $d$  at time  $t$  being a party’s first hunting trip that day and  $g_{s,d}(u)$  denotes the expected party size. The intuition behind this formula is to only count a party during their first hunting trip in the survey region, and not count subsequent hunting trips to other sites on the same day.

The estimated number of recreationists in the Southwest region over time is shown in Figure 4. Peak turnout is on the pheasant and quail opening weekend (last weekend

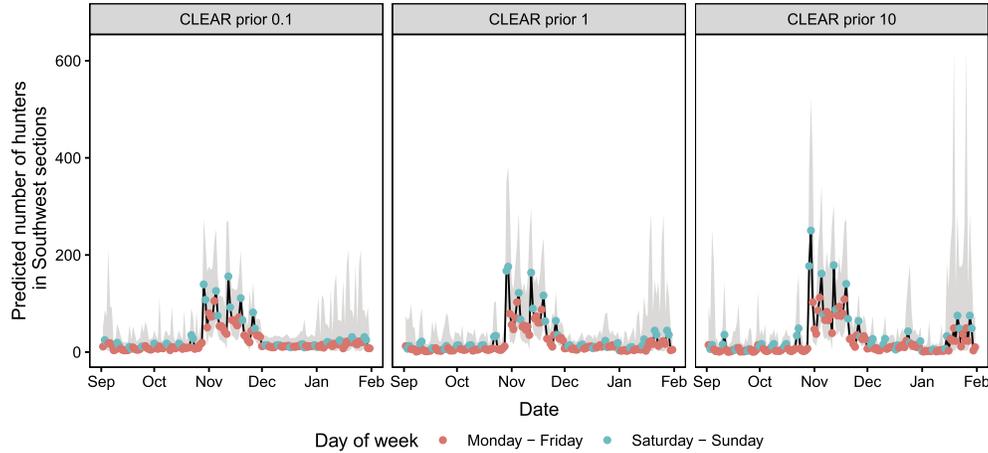


Figure 4: Estimated fall 2016 recreationist population across all surveyed public access sites in the Southwest region by day; the shaded area shows the 95% credible interval; the black line shows the median estimates.

in October), and turnout remains elevated throughout the pheasant, quail and deer hunting seasons. Our estimates suggest that the typical daily Southwest Nebraska hunter population during the four peak weekends (pheasant and quail opening weekend, and the three following weekends) is about 121 recreationists. Additional estimates are in Table 3.

	Prior 0.1	Prior 1	Prior 10
Recreationists	3,748	3,445	4,042
Recreationists to private sites	1,303	1,1157	1,403
Recreationists to state sites	2,445	2,288	2,640
Recreationists on weekends	1,608	1,659	1,992
Recreationists Monday-Friday	2,139	1,787	2,051
Recreationists to private sites on pheasant opening weekend	82	104	117
Recreationists to state sites on deer opening weekend	127	135	147
Recreationists during peak season (incl. weekends)	1,965	2,031	2,405
Recreationists on 4 peak Saturdays	533	570	658
Recreationists on 4 peak Sundays	341	396	484

Table 3: Estimated hunter population (posterior medians) during the fall 2016 season across all 89 surveyed sites in Southwest Nebraska. Figures are the sum over the daily population estimates for all 89 surveyed sites.

### 4.5 Potential implications for policy and land management

By conducting CLEAR analysis to assess the spatial and temporal distribution of different hunter typologies at large spatial scales, we provide a first high-resolution site use

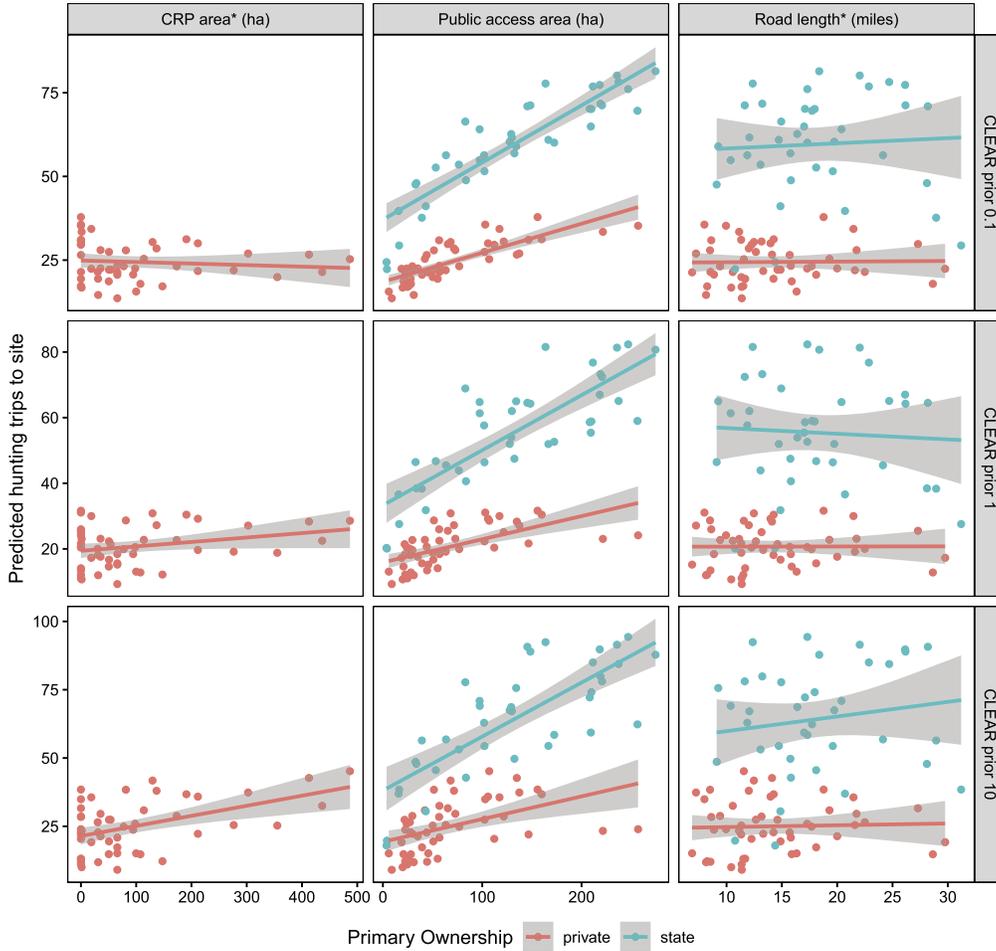


Figure 5: Points show estimated hunting trips (based on posterior medians) made to individual sites arranged by selected covariate values. Covariates marked with an asterisk indicate values of the section of the hunting site and surrounding sections. Lines are ad-hoc linear regression lines through the points to aid in visual interpretation.

and hunter population estimate for public access hunting grounds. Figure 5 shows the estimated site use for the surveyed sites along several selected covariates. The major visible effect is the size of the public access hunting site—the bigger the site, the more trips are made to it. Primary ownership appears to play a role, but that effect is most certainly not a causal one, but rather one of correlation of the types of sites that are state or privately owned and the prevalent target species at these sites (see the seasonal pattern visible in Figure 3). Road density in a section does not appear to affect site use substantially, while a positive effect of CRP on privately owned sites is visible in the models with weaker priors.

High temporal resolution estimates of site use can be of great use to public lands managers managing public access land portfolios. However, no previous such studies exist, given the lack of existing methodology to obtain high-resolution estimates of immigration (hunting trips) from the collected data. CLEAR fills this need. Potential future application of CLEAR includes wildlife agencies using site use estimates to predict the likely economic contribution of public lands to rural economies when negotiating new land purchases and private-public hunting initiatives with rural communities.

## 5 Concluding remarks

Our results regarding the convolution of the binomial and Poisson distributions provide a theoretical foundation for elimination of computationally expensive modeling of latent states to estimate the effects of covariates on population immigration and emigration. Furthermore, by linking a differential equation describing the population mean effects to the appropriate likelihoods for immigration and emigration, we provide a theory-based justification for current practice in ecological studies combining the Poisson and binomial distributions to model these effects, and enable estimates and predictions of unobserved quantities that can be derived from analysis of the underlying differential equation.

CLEAR may also prove useful for ecological investigations of populations whose abundance changes rapidly in space and time. For example, migratory bird populations may use a given stopover for a relatively brief time, but every stop on a migration route is critical. Modeling migration traffic and stopover use of migratory populations at fine temporal resolution may help to predict which stopover sites will become critical at what times, facilitating conservation of habitat that becomes extremely important for a short time. Likewise, CLEAR could also help model the spatial and temporal distribution of populations using ephemeral habitat, such as breeding amphibians, who likewise have very specific habitat needs for a specific time frame.

Predicting spatial variation in population dynamics is an objective of pressing importance for ecological inference (Guisan and Thuiller, 2005). Ecological studies often create multiple datasets on the same population for the same time interval (for example, camera traps and hair snares, or radio telemetry and point counts), but with varying, and possibly inhomogenous, temporal resolutions. Integrating such datasets into unified, spatially-explicit population models will empower ecologists to ask nuanced questions about how spatial variation in resources and constraints affects not only the distribution, but the persistence of populations and communities.

We expect that the methods presented in this manuscript will influence future analysis of high-frequency count data in ecology and other fields. Future methodological work could include the extension of our regression analysis to a full-fledged Bayesian ecological abundance time series model for count data and the development of online learning methods. Methodology for online learning could be based on variational concepts such as those used for high-dimensional real-valued time series (Gruber and West, 2016, 2017), or particle filtering as recently used for multivariate time series for count data (Aktekin et al., 2018).

## Supplementary Material

Supplementary Material for “Estimating the Use of Public Lands: Integrated Modeling of Open Populations with Convolution Likelihood Ecological Abundance Regression” (DOI: [10.1214/19-BA1152SUPP](https://doi.org/10.1214/19-BA1152SUPP); .pdf).

## References

- Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2010). “Estimation of immigration rate using integrated population models.” *Journal of Applied Ecology*, 47(2): 393–400. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2664.2010.01789.x> 1177
- Aktekin, T., Polson, N., and Soyer, R. (2018). “Sequential Bayesian Analysis of Multivariate Count Data.” *Bayesian Anal.*, 13(2): 385–409. MR3780428. doi: <https://doi.org/10.1214/17-BA1054>. 1195
- Anderson, W. L. and David, L. M. (1998). “Results of the 1996-97 Illinois Pheasant Hunter Survey.” report, Illinois Department of Natural Resources, Division of Wildlife Resources, Office of Resource Conservation, Federal Aid Coordinator. 1190
- Barker, R. J., Schofield, M. R., Link, W. A., and Sauer, J. R. (2018). “On the reliability of N-mixture models for count data.” *Biometrics*, 74(1): 369–377. MR3777957. doi: <https://doi.org/10.1111/biom.12734>. 1183
- Brauer, F. and Castillo-Chavez, C. (2001). *Mathematical Models in Population Biology and Epidemiology*. Springer, New York, NY. MR1822695. doi: <https://doi.org/10.1007/978-1-4757-3516-1>. 1173
- Brinley Buckley, E. M., Allen, C. R., Forsberg, M., Farrell, M., and Caven, A. J. (2017). “Capturing change: the duality of time-lapse imagery to acquire data and depict ecological dynamics.” *Ecology and Society*, 22(3). 1174
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software, Articles*, 76(1): 1–32. <https://www.jstatsoft.org/v076/i01> 1189
- Chandler, R. B., Muths, E., Sigafus, B. H., Schwalbe, C. R., Jarchow, C. J., and Hossack, B. R. (2015). “Spatial occupancy models for predicting metapopulation dynamics and viability following reintroduction.” *Journal of Applied Ecology*, 52(5): 1325–1333. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2664.12481> 1174
- Dail, D. and Madsen, L. (2011). “Models for Estimating Abundance from Repeated Counts of an Open Metapopulation.” *Biometrics*, 67(2): 577–587. MR2829026. doi: <https://doi.org/10.1111/j.1541-0420.2010.01465.x>. 1183
- Greenberg, S. (2015). “Timelapse2.” <http://saul.cpsc.ucalgary.ca/timelapse/pmwiki.php> 1185

- Gruber, L. F., Stuber, E. F., Wszola, L. S., and Fontaine, J. J. (2019). “Supplementary Material for “Estimating the Use of Public Lands: Integrated Modeling of Open Populations with Convolution Likelihood Ecological Abundance Regression”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/19-BA1152>. 1183, 1190
- Gruber, L. F. and West, M. (2016). “GPU-Accelerated Bayesian Learning and Forecasting in Simultaneous Graphical Dynamic Linear Models.” *Bayesian Anal.*, 11(1): 125–149. MR3447094. doi: <https://doi.org/10.1214/15-BA946>. 1195
- Gruber, L. F. and West, M. (2017). “Bayesian online variable selection and scalable multivariate volatility forecasting in simultaneous graphical dynamic linear models.” *Econometrics and Statistics*, 3: 3–22. MR3666239. doi: <https://doi.org/10.1016/j.ecosta.2017.03.003>. 1195
- Guisan, A. and Thuiller, W. (2005). “Predicting species distribution: offering more than simple habitat models.” *Ecology Letters*, 10. 1195
- Hanski, I. and Simberloff, D. (1997). “The Metapopulation Approach, Its History, Conceptual Domain, and Application to Conservation.” In Hanski, I. and Gilpin, M. E. (eds.), *Metapopulation Biology*, 5–26. San Diego: Academic Press. <https://www.sciencedirect.com/science/article/pii/B9780123234452500031> 1173
- Hiller, T. L., Taylor, J. S., Lusk, J. J., Powell, L. A., and Tyre, A. J. (2015). “Evidence that the conservation reserve program slowed population declines of pheasants on a changing landscape in Nebraska, USA: Pheasant Populations and the CRP.” *Wildlife Society Bulletin*, 39. 1190
- Huber, N., Kéry, M., and Pasinelli, G. (2017). “Occupancy dynamics of the Wood Warbler *Phylloscopus sibilatrix* assessed with habitat and remote sensing data.” *Ibis*, 159(3): 623–637. <https://onlinelibrary.wiley.com/doi/abs/10.1111/ibi.12472> 1174
- Kéry, M. (2018). “Identifiability in N-mixture models: a large-scale screening test with bird data.” *Ecology*, 99(2): 281–288. <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2093> 1183
- Kéry, M. and Schmid, H. (2006). “Estimating species richness: calibrating a large avian monitoring programme.” *Journal of Applied Ecology*, 43(1): 101–110. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2664.2005.01111.x> 1174
- Knape, J. and Korner-Nievergelt, F. (2016). “On assumptions behind estimates of abundance from counts at multiple sites.” *Methods in Ecology and Evolution*, 7(2): 206–209. <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12507> 1183
- Link, W. A., Schofield, M. R., Barker, R. J., and Sauer, J. R. (2018). “On the robustness of N-mixture models.” *Ecology*, 99(7): 1547–1551. <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2362> 1183
- Lone, K., Loe, L. E., Gobakken, T., Linnell, J. D. C., Odden, J., Remmen, J., and Mysterud, A. (2014). “Living and dying in a multi-predator landscape of fear: roe deer

- are squeezed by contrasting pattern of predation risk imposed by lynx and humans.” *Oikos*, 123. 1184
- Mckenzie, E. (1988). “Some ARMA models for dependent sequences of poisson counts.” *Advances in Applied Probability*, 20(4): 822–835. MR0968000. doi: <https://doi.org/10.2307/1427362>. 1174, 1179
- Mecozzi, G. E. and Guthery, F. S. (2008). “Behavior of Walk-Hunters and Pointing Dogs During Northern Bobwhite Hunts.” *Journal of Wildlife Management*. <http://www.bioone.org/doi/abs/10.2193/2007-555> 1184
- Morin, D. J., Fuller, A. K., Royle, J. A., and Sutherland, C. (2017). “Model-based estimators of density and connectivity to inform conservation of spatially structured populations.” *Ecosphere*, 8(1): e01623. <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecs2.1623> 1174
- Nebraska Game & Parks Commission (2018-05-09a). “Nebraska Public Access Atlas.” <https://maps.outdoornebraska.gov/PublicAccessAtlas/> 1175, 1184
- Nebraska Game & Parks Commission (2018-05-09b). “Upland Hunting Guide.” <http://outdoornebraska.gov/upland/> 1184
- Royle, J. A. (2004). “N-Mixture Models for Estimating Population Size from Spatially Replicated Counts.” *Biometrics*, 60(1): 108–115. MR2043625. doi: <https://doi.org/10.1111/j.0006-341X.2004.00142.x>. 1183
- Ryan, E. L. and Shaw, B. (2011). “Improving Hunter Recruitment and Retention.” *Human Dimensions of Wildlife*, 16(5). 1184
- Schaub, M. and Abadi, F. (2011). “Integrated population models: a novel analysis framework for deeper insights into population dynamics.” *Journal of Ornithology*, 152(1): 227–237. <https://doi.org/10.1007/s10336-010-0632-7> 1177
- Schoen, R. (1988). “Practical Uses of Multistate Population Models.” *Annual Review of Sociology*, 14(1): 341–361. PMID: 12315442. <https://doi.org/10.1146/annurev.so.14.080188.002013> 1173
- Sih, A., Ferrari, M. C. O., and Harris, D. J. (2011). “Evolution and behavioural responses to human-induced rapid environmental change.” *Evolutionary Applications*, 4(2): 367–387. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-4571.2010.00166.x> 1174
- Stedman, R., Diefenbach, D. R., Swope, C. B., Finley, J. C., Luloff, A. E., Zinn, H. C., San Julian, G. J., and Wang, G. A. (2004). “Integrating wildlife and human-dimensions research methods to study hunters.” *Journal of Wildlife Management*. [http://www.bioone.org/doi/abs/10.2193/0022-541X\(2004\)068\[0762:IWAHRM\]2.0.CO;2](http://www.bioone.org/doi/abs/10.2193/0022-541X(2004)068[0762:IWAHRM]2.0.CO;2) 1175
- Steutel, F. W. and van Harn, K. (1979). “Discrete Analogues of Self-Decomposability and Stability.” *Ann. Probab.*, 7(5): 893–899. <https://doi.org/10.1214/aop/1176994950> MR0542141. 1179

- Stutel, F. W., Vervaat, W., and Wolfe, S. J. (1983). “Integer-valued branching processes with immigration.” *Advances in Applied Probability*, 15(4): 713–725. MR0721702. doi: <https://doi.org/10.2307/1427320>. 1174, 1179
- Stuber, E. F., Gruber, L. F., and Fontaine, J. J. (2017). “A Bayesian method for assessing multi-scale species-habitat relationships.” *Landscape Ecology*, 32(12): 2365–2381. <https://doi.org/10.1007/s10980-017-0575-y> 1181
- Stuber, E. F., Gruber, L. F., and Fontaine, J. J. (2018). “Predicting species-habitat relationships: does body size matter?” *Landscape Ecology*. <https://doi.org/10.1007/s10980-018-0648-6> 1174
- U.S. Fish and Wildlife Service (2016). “2016 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation.” <https://www2.census.gov/programs-surveys/fhwar/publications/2016/fhw16-nat.pdf> 1175
- U.S. Geological Survey (2018). “The Public Land Survey System (PLSS).” [https://nationalmap.gov/small\\_scale/a\\_plss.html](https://nationalmap.gov/small_scale/a_plss.html) 1184
- Zipkin, E. F., Rossman, S., Yackulic, C. B., Wiens, J. D., Thorson, J. T., Davis, R. J., and Grant, E. H. C. (2017). “Integrating count and detection–nondetection data to model population dynamics.” *Ecology*, 98(6): 1640–1650. <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.1831> 1174, 1177
- Zipkin, E. F., Thorson, J. T., See, K., Lynch, H. J., Grant, E. H. C., Kanno, Y., Chandler, R. B., Letcher, B. H., and Royle, J. A. (2014). “Modeling structured population dynamics using data from unmarked individuals.” *Ecology*, 95(1): 22–29. <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/13-1131.1> 1177, 1186

### Acknowledgments

Funding for this project was received from Federal Aid in Wildlife Restoration projects W-115-R, administered by the Nebraska Game and Parks Commission. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The Nebraska Cooperative Fish and Wildlife Research Unit is supported by a cooperative agreement among the U.S. Geological Survey, the Nebraska Game and Parks Commission, the University of Nebraska, the U.S. Fish and Wildlife Service, and the Wildlife Management Institute.

We would like to thank C. Hellmayr and two anonymous reviewers for their valuable comments on earlier versions of this manuscript.