Running head: Pinpointing spatial features

Leveraging constraints and biotelemetry data to pinpoint repetitively used spatial features

Brian M. Brost[1]

Department of Fish, Wildlife, and Conservation Biology, Colorado State University

Mevin B. Hooten

U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit

Department of Fish, Wildlife, and Conservation Biology, Colorado State University

Department of Statistics, Colorado State University

Robert J. Small

Division of Wildlife Conservation, Alaska Department of Fish and Game

Abstract. Satellite telemetry devices collect valuable information concerning the sites visited by animals, including the location of central places like dens, nests, rookeries, or haul-outs. Existing methods for estimating the location of central places from telemetry data require user-specified thresholds and ignore common nuances like measurement error. We present a fully model-based approach for locating central places from telemetry data that accounts for multiple sources of uncertainty and uses all of the available locational data. Our general framework consists of an observation model to account for large telemetry measurement error and animal movement, and a highly flexible mixture model specified using a Dirichlet process to identify the location of central places. We also quantify temporal patterns in central place use by incorporating ancillary behavioral data into the model; however, our framework is also suitable when no such behavioral data exist. We apply the model to a simulated data set as proof of concept. We then illustrate

[1]Corresponding author; E-mail: bmbrost@gmail.com.

our framework by analyzing an Argos satellite telemetry data set on harbor seals (*Phoca vitulina*) in the Gulf of Alaska, a species that exhibits fidelity to terrestrial haul-out sites.

Key words: Harbor seal, Phoca vitulina, haul-out, Dirichlet process, mixture model, Bayesian analysis, hierarchical model, nonparametric, basis function, temporal dependence, integrated data model, data fusion.

Introduction

Many animal species return regularly to one or more central places like a den, nest, roost, or foraging site. Central places can be located by sighting individuals during aerial (Montgomery et al. 2007) or ground-based surveys (Blakesley et al. 1992), or by using radio-telemetry equipment to locate individuals in the field (Holloran and Anderson 2005); however, direct observation may only provide a snapshot of the animal's behavior if surveys are infrequent (Ruprecht et al. 2012), and could be altogether impractical when surveys are encumbered by remote locations, rugged terrain, or otherwise difficult conditions. We address these issues using a model-based approach for locating central places from satellite telemetry data.

Satellite telemetry devices collect regular sequences of animal locations (Tomkiewicz et al. 2010), data that contain valuable information concerning the sites visited over a monitoring period. Repeated use of a site often yields multiple telemetry locations collected at that site. Therefore, clusters of locations in mapped telemetry data are important indicators of a central place (Knopff et al. 2009).

When deviations between true animal locations and the observed telemetry locations are small (i.e., small telemetry measurement error), clusters are well-defined. Accordingly, central places can be located by identifying clusters consisting of some pre-specified number of telemetry locations collected within a certain distance and time frame (Anderson and Lindzey 2003, Knopff et al. 2009). However, results are sensitive to the distance and time thresholds used (Zimmermann et al. 2007). Moreover, distance thresholds fail when telemetry measurement error is large. Large errors lead to diffuse clusters, which, in turn, create uncertainty in the location of a central place as well as the composition of the clusters themselves. For example, observed telemetry locations can plausibly originate from more than one central place (i.e., cluster membership is ambiguous), or locations collected at a central place can be confused with locations collected during movements away from the site. Therefore, a method that accounts for telemetry measurement error is required.

We present a model-based approach for estimating the location of central places from satellite telemetry data. Our approach incorporates an observation model that explicitly accounts for measurement error, and uses a mixture model as a device for exposing latent structure (i.e., clustering) in telemetry location data. The mixture model is specified using a flexible Dirichlet process prior, a well-developed Bayesian nonparametric model that adapts its complexity to the data at hand. We also quantify temporal patterns in central place use (i.e., factors affecting when a central place is used) by incorporating ancillary data related to animal behavior into the model; however, we also extend the model to situations when no such behavioral data exist. We first apply the model to a simulated data set as proof of concept. We then illustrate our framework using an Argos satellite telemetry data set on harbor seals (*Phoca vitulina*) in the Gulf of Alaska. Harbor seals are central place foragers that exhibit fidelity to terrestrial haul-out sites (Lowry et al. 2001).

Telemetry Data

The model we propose can be applied to various telemetry data types like VHF, GPS, or geolocation telemetry. We focus on Argos satellite telemetry data like those in our harbor seal data set that were calculated via the Argos least-squares positioning algorithm (Service Argos 2015). These data require special treatment because they exhibit an x-shaped error distribution that has greatest error variance along the NW-SE and NE-SW axes, a consequence of the polar orbiting Argos satellites and error that is largest in the direction perpendicular to the orbit (Costa et al. 2010, Douglas et al. 2012). Furthermore, valid Argos telemetry locations are assigned one of six location classes (3, 2, 1, 0, A, and B), each of which exhibits different error patterns and magnitudes.

In addition to positional data, modern telemetry devices often collect ancillary data related to animal behavior (Tomkiewicz et al. 2010) that can be helpful for partitioning when individuals are actively using a central place versus other resources. The harbor seals in our data set, for example, were equipped with satellite-linked depth recorders that gathered information pertaining to diving behavior. Specifically, we use information from an on-board conductivity sensor that differentiates when a tag is wet (low resistance) versus dry (high resistance) as a surrogate for central place use. Resistance values ranged from 0-255, which we converted into a binary indicator for haul-out status using a threshold value of 127 (i.e., resistance values > 127 were categorized as hauled-out). The devices were programmed with a delay (10 consecutive readings at 45 sec. intervals) to prevent spurious wet/dry state transitions associated with splashing on a haul-out or short dry periods experienced by the sensor while a seal was surfaced but swimming; therefore, these wet/dry data reliably indicate when an individual is hauled-out on shore (dry) or at-sea (wet).

Model Formulation

Let $\mathbf{s}(t) \equiv (s_x(t), s_y(t))'$ represent the pair of coordinates for an observed telemetry location at time $t \in \mathcal{T}$, and $\boldsymbol{\mu}(t) \equiv (\mu_x(t), \mu_y(t))'$ represent the coordinates for a corresponding latent central place. We denote the spatial support of central places as $\widetilde{\mathcal{S}}$ and the ancillary behavioral data as $y(t)$. In the case of harbor seals, $\widetilde{\mathcal{S}}$ represents the coastline where haul-out sites can occur and $y(t) \in \{0, 1\}$, where 0 indicates the individual is at-sea and 1 indicates the individual is on-shore using terrestrial resources.

Observation model.—The observed telemetry locations arise from a process that reflects animal movement and measurement error. Movement influences the true animal locations which are then observed imperfectly due to the telemetry measurement process. We accommodate various error patterns using a flexible mixture distribution, which itself is conditioned on the ancillary behavioral data to accommodate movement. First, consider a model for telemetry locations collected while the individual is at a central place (i.e., $y(t) = 1$):

$$\mathbf{s}(t) \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma}), & \text{with prob. } p(t) \\ \mathcal{N}(\boldsymbol{\mu}(t), \widetilde{\boldsymbol{\Sigma}}), & \text{with prob. } 1 - p(t). \end{cases} \quad (1)$$

In Eq. 1, an observed telemetry location ($\mathbf{s}(t)$) arises from a mixture of multivariate normal distributions with mean $\boldsymbol{\mu}(t)$ corresponding to the location of a central place, and variance-covariance matrices $\boldsymbol{\Sigma}$ or $\widetilde{\boldsymbol{\Sigma}}$ that describe telemetry measurement error. The matrix $\boldsymbol{\Sigma}$ is parameterized in a flexible manner (Brost et al. 2015, Buderman et al. 2016):

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho\sqrt{a} \\ \rho\sqrt{a} & a \end{bmatrix}, \quad (2)$$

where σ^2 quantifies measurement error in the longitude direction, a modifies σ^2 to describe error in the latitude direction, and ρ describes the correlation between errors in the two directions. The matrix $\widetilde{\boldsymbol{\Sigma}}$ equals $\boldsymbol{\Sigma}$ on the diagonal, but the off-diagonal elements are $-\rho\sqrt{a}$. This model specification accounts for circular ($a = 1$) and elliptical ($a \neq 1$) errors when $\rho = 0$, as well as x-shaped error patterns evident in Argos telemetry data when $\rho \neq 0$.

We model telemetry locations collected while the individual is not at the central place (i.e., $y(t) = 0$) in a fashion similar to Eq. 1:

$$\mathbf{s}(t) \sim \begin{cases} \mathcal{N}(\boldsymbol{\mu}(t), \boldsymbol{\Sigma} + \sigma_\mu^2 \mathbf{I}), & \text{with prob. } p(t) \\ \mathcal{N}(\boldsymbol{\mu}(t), \widetilde{\boldsymbol{\Sigma}} + \sigma_\mu^2 \mathbf{I}), & \text{with prob. } 1 - p(t), \end{cases} \quad (3)$$

except the variance-covariance structure in Eq. 3 is augmented by σ_μ^2, a parameter accounting for dispersion due to animal movement about the central place. In other words,

$\boldsymbol{\mu}(t)$ and σ_μ^2 define the center and spread of an individual's "homerange." As in Eq. 1, $\boldsymbol{\Sigma}$ and $\widetilde{\boldsymbol{\Sigma}}$ account for error in the telemetry measurement process.

The observation model in Eq. 3 represents an integrated likelihood (Berger et al. 1999). Consider, for example, the hierarchical model

$$\mathbf{s}(t) \sim \mathcal{N}\left(\tilde{\boldsymbol{\mu}}(t), \sigma^2\mathbf{I}\right) \tag{4}$$

$$\tilde{\boldsymbol{\mu}}(t) \sim \mathcal{N}\left(\boldsymbol{\mu}(t), \sigma_\mu^2\mathbf{I}\right), \tag{5}$$

where $\tilde{\boldsymbol{\mu}}(t)$ is the true but unobserved animal location. The parameters $\boldsymbol{\mu}(t)$, σ^2, and σ_μ^2 are defined as in Eqs. 1-3, but note that the telemetry error structure in Eq. 4 is simplified for the purposes of illustration. In principle, we could estimate the true location $\tilde{\boldsymbol{\mu}}(t)$; however, our interest here is not the true locations but rather the location of the central place, $\boldsymbol{\mu}(t)$. Therefore, we treat $\tilde{\boldsymbol{\mu}}(t)$ as a "nuisance" parameter and remove it from the likelihood by integration (i.e., Rao-Blackwellization; Berger et al. 1999):

$$\int_{\tilde{\boldsymbol{\mu}}(t)} \mathcal{N}\left(\mathbf{s}(t) \mid \tilde{\boldsymbol{\mu}}(t), \sigma^2\mathbf{I}\right) \mathcal{N}\left(\tilde{\boldsymbol{\mu}}(t) \mid \boldsymbol{\mu}(t), \sigma_\mu^2\mathbf{I}\right) d\tilde{\boldsymbol{\mu}}(t) = \mathcal{N}\left(\mathbf{s}(t) \mid \boldsymbol{\mu}(t), \sigma^2\mathbf{I} + \sigma_\mu^2\mathbf{I}\right). \tag{6}$$

Aside from the simplified error structure, the resulting marginal distribution is the same as Eq. 3 and has a reduced parameter space compared to Eqs. 4 and 5. It also yields a Markov chain Monte Carlo (MCMC) algorithm that is typically quicker to converge (Finley et al. 2015). Models for animal movement where individuals are attracted to a particular point are also available if inference concerning $\tilde{\boldsymbol{\mu}}(t)$ is desired (Blackwell 2003, McClintock et al. 2012); however, these methods require the number of central places used by an individual to be known.

We define $p(t) = 0.5$ because the orbital plane of Argos satellites changes continuously and observations are equally likely to arise from either mixture component. The parameters related to measurement error (i.e., σ^2, ρ, and a) are estimated for different Argos location quality classes (Appendix S1). Alternatively, Eq. 2 can be adapted to accommodate a continuous metric of location quality (e.g., GPS dilution of precision) or the Argos satellite telemetry location error ellipse (McClintock et al. 2014).

Spatial process model.—As specified in the observation model (Eqs. 1 and 3), a telemetry location arises from an unknown (but estimable) central place, $\boldsymbol{\mu}(t)$. When considering multiple telemetry locations recorded over some period of time, the number of unique central places used by an individual is potentially > 1, but the exact number is unknown. Modeling central places is further complicated by possible multimodality (central places located in disjoint areas) and skewness (some central places are close together). We resolve these issues (i.e., multimodality, skewness, and an unknown number of central places) by using a Dirichlet process, a widely used probability model for unknown distributions that exhibits an important clustering property (Ferguson 1973, Hjort 2010). Following the constructive, stick-breaking representation of a Dirichlet process (Sethuraman 1994, Ishwaran and James 2001), we model $\boldsymbol{\mu}(t)$ as a mixture of infinitely many components:

$$\boldsymbol{\mu}(t) \sim \sum_{j=1}^{\infty} \pi_j \delta_{\mu_j}, \tag{7}$$

where $\boldsymbol{\mu}_j$ is the location of a potential central place, δ_{μ_j} is a point mass (or "atom") at $\boldsymbol{\mu}_j$, π_j is the corresponding mixing proportion, and $\sum_{j=1}^{\infty} \pi_j = 1$. Because Eq. 7 is a discrete distribution, draws from it are generally not distinct, thereby inducing replication in the values for $\boldsymbol{\mu}(t)$. Thus, realizations from the Dirichlet process simultaneously provide a value for $\boldsymbol{\mu}(t)$ and partition telemetry locations with the same value for $\boldsymbol{\mu}(t)$ into clusters. The distinction between $\boldsymbol{\mu}_j$ and $\boldsymbol{\mu}(t)$ is subtle. The $\boldsymbol{\mu}_j$, for $j = 1, \ldots, \infty$, are unique and represent the location of potential central places. The $\boldsymbol{\mu}(t)$, on the other hand, have a functional interpretation because they are time-specific and associate a $\boldsymbol{\mu}_j$ to each telemetry location $\mathbf{s}(t)$. Greater replication of $\boldsymbol{\mu}(t)$, for $t \in \mathcal{T}$, confers higher intensity use of the associated central place (i.e., more telemetry locations associated with the same central place). Note that, even though the Dirichlet process assumes infinitely many mixture components (central places), only a finite number are used to generate the observed data.

We formulate π_j using a stick-breaking process (Sethuraman 1994):

$$\pi_j = \eta_j \prod_{l<j} (1 - \eta_l), \tag{8}$$

where $\eta_j \sim \text{Beta}(1, \theta)$ and θ is a concentration parameter that controls the prior expected number of mixture components in the Dirichlet process. To describe the stick-breaking process, begin with a stick of unit length that represents the total probability allocated to the infinitely many mixture components in Eq. 7. Initially, we break off a piece of length $\eta_1 \sim \text{Beta}(1, \theta)$ from the stick and assign this probability ($\pi_1 = \eta_1$) to the first component, $\boldsymbol{\mu}_1$. Next, we break off another proportion $\eta_2 \sim \text{Beta}(1, \theta)$ from the remaining length of stick $(1 - \eta_1)$ and assign this probability ($\pi_2 = \eta_2 (1 - \eta_1)$) to the second component, $\boldsymbol{\mu}_2$. As the process is repeated, the stick gets shorter such that the lengths (i.e., mixing proportions) assigned to components with a higher index decrease stochastically. The concentration parameter (θ) controls the rate of decrease.

In practice, we implement the Dirichlet process using a truncation approximation (Ishwaran and James 2001). For a sufficiently high index J, notice that $\sum_{J+1}^{\infty} \pi_j \approx 0$ because the mixing proportions decrease in the index j. Thus, an accurate approximation to the infinite Dirichlet process (Eq. 7) can be obtained by letting $\eta_J = 1$, resulting in $\pi_j = 0$ for $j = J+1, \ldots, \infty$. The index J is an upper bound on the number of mixture components in Eq. 7, not the number of components necessary to model the observed data.

Temporal process model.—We model the ancillary behavioral data using a binary probit regression formulated under a data augmentation approach (Albert and Chib 1993, Johnson et al. 2012, Dorazio and Rodriguez 2012). In particular, we introduce the parameter $v(t)$ as a continuous, latent version of the binary process $y(t)$, which we model as a normal random variable with unit variance:

$$v(t) \sim \mathcal{N}\left(\mathbf{x}(t)' \boldsymbol{\beta} + \mathbf{w}(t)' \boldsymbol{\alpha}, 1\right). \tag{9}$$

This expression represents a semiparametric regression with mean structure that includes parametric and nonparametric components (Hastie et al. 2009, Ruppert et al. 2003). The parametric component consists of a vector of time-varying covariates that affect the probability of central place use, $\mathbf{x}(t)$, and a corresponding vector of coefficients, $\boldsymbol{\beta}$. The nonparametric component, $\mathbf{w}(t)'\boldsymbol{\alpha}$, is described below. Assuming $y(t) = 1$ if $v(t) > 0$ and $y(t) = 0$ if $v(t) \leq 0$, the specification in Eq. 9 implies the probit regression model

$$y(t) \sim \text{Bernoulli}\left(\Phi\left(\mathbf{x}(t)'\boldsymbol{\beta} + \mathbf{w}(t)'\boldsymbol{\alpha}\right)\right), \tag{10}$$

where Φ is the standard normal cumulative distribution function. The auxiliary variable specification in Eqs. 9 and 10 streamlines computation because the associated full-conditional distributions are known and can be sampled in closed form when fitting the model using MCMC.

We use the nonparametric component of Eq. 9 to account for temporal autocorrelation, which often occurs in data collected over time from a single individual (e.g., $y(t)$). The nonparametric component consists of a linear combination of basis functions evaluated at time t, $\mathbf{w}(t)$, and the vector of basis coefficients, $\boldsymbol{\alpha}$ (Ruppert et al. 2003). The coefficients weight the basis functions to produce a smooth process through time, thereby inducing dependence among observations. The basis functions are arbitrary and should have features that match those of the underlying process being estimated. Commonly used basis functions include splines, wavelets, and Fourier series. The number of functions should also reflect the temporal resolution of that process (Ruppert et al. 2003).

Prior distributions.—To complete the Bayesian formulation of this model, we specify prior distributions for unknown parameters. We assume $\boldsymbol{\beta} \sim \text{N}\left(\boldsymbol{\mu}_{\beta}, \sigma^2_{\beta}\mathbf{I}\right)$, $\theta \sim \text{Gamma}(r_{\theta}, q_{\theta})$, $\log(\sigma_{\mu}) \sim \mathcal{N}(\mu_{\sigma}, \sigma^2_{\sigma})$, and $\sigma \sim \text{Uniform}(0, u)$, with similar uniform priors for ρ and a. The lognormal distribution for σ_{μ} allows prior information concerning animal movement and homerange size, if available, to be incorporated into the model. We adopt a penalized approach to avoid overfitting $\boldsymbol{\alpha}$ by assuming $\boldsymbol{\alpha} \sim \mathrm{N}\left(\mathbf{0}, \sigma_{\alpha}^{2}\mathbf{I}\right)$ and $\sigma_{\alpha}^{2} \sim \mathrm{IG}\left(r_{\alpha}, q_{\alpha}\right)$ (Ruppert et al. 2003). The prior for $\boldsymbol{\mu}_j$, referred to as the base distribution of the Dirichlet process (Hjort 2010), determines where the atoms $\delta_{\boldsymbol{\mu}_j}$ tend to be located. We assume $\boldsymbol{\mu}_j \sim f_{\widetilde{\mathcal{S}}}\left(\mathbf{S}\right)$, where $\mathbf{S}$ is a matrix containing all of the observed telemetry locations and $f_{\widetilde{\mathcal{S}}}\left(\mathbf{S}\right)$ represents the density of telemetry locations in $\widetilde{\mathcal{S}}$. We approximate $f_{\widetilde{\mathcal{S}}}\left(\mathbf{S}\right)$ using a kernel density estimator evaluated over a rasterized domain $\widetilde{\mathcal{S}}$. See Appendix S1 for the full model specification and Appendix S2 for details regarding model implementation.

Model Application

Simulated data example

We demonstrate our modeling framework when parameters are known in a simulated data example. Figure 1 shows 1,000 locations simulated from the model using parameters obtained from an analysis of harbor seal telemetry data (see Case study below). To simplify presentation of results, simulated locations were randomly allocated to Argos location classes 3, 0, and B (high-, medium-, and low-accuracy locations). We set $J = 50$ in the truncation approximation to the Dirichlet process and modeled dependence in central place use with B-spline basis functions ($\mathbf{w}\left(t\right)$). B-splines are commonly used in semiparametric regression because they have local support and stable numerical properties (Ruppert et al. 2003). We fit the model using a MCMC algorithm written in R (provided in Data S1; R Development Core Team 2015).

Inference concerning $\boldsymbol{\mu}\left(t\right)$, the spatial intensity of central place use, is summarized in Figure 1. Posterior probability is concentrated near known central places, and inference is more certain for central places associated with many telemetry locations (i.e., locations that were heavily used). Posterior probability for $\boldsymbol{\mu}_j$, the location of potential central places, is more diffuse than that of $\boldsymbol{\mu}\left(t\right)$, but still generally concentrated near central places (Appendix S3). The model recovers parameters related to telemetry measurement error, animal movement, and the temporal process of central place use (Appendix S3). Additional simulated data examples are presented in Appendix S4.

Case study: Harbor seals

To demonstrate our approach with real data, we apply our model to Argos satellite telemetry locations collected from a harbor seal near Kodiak Island, Alaska (Fig. 2). Harbor seals repeatedly use terrestrial haul-out sites along the coastline ($\widetilde{\mathcal{S}}$), which we represented using a 100-m resolution raster. Haul-out behavior changes over time due to physiological functions (thermoregulation, molting, pupping, etc.) and environmental conditions (e.g., tidal state) that affect the availability of haul-out sites (London et al. 2012). Thus, we evaluated the affect of several temporal covariates on the use of haul-out sites: the number of hours since solar noon (13:00 hours), the number of hours since low tide, and the number of days since August 15 and its quadratic effect. Tide information was obtained from the nearest National Oceanic and Atmospheric Administration station (Kodiak Island, ID: 9457292). We set $J = 50$ in the truncation approximation to the Dirichlet process, which greatly exceeds the expected number of haul-out sites used by a single harbor seal. We modeled the temporal haul-out process using B-splines ($\mathbf{w}(t)$) defined at 6-hour intervals. In addition to allowing for smooth patterns in the probability of haul-out use, a basis expansion defined at this interval allows haul-out behavior to vary throughout day.

Inference concerning the intensity of haul-out site use ($\boldsymbol{\mu}(t)$) is shown in Figure 2. Posterior probability is concentrated in three regions, generally occurring near clustered telemetry locations. The highest posterior probability occurs along the northernmost coastline of Ugak Bay, indicating this area was most actively used by the individual. Similar to the simulated data example, inference concerning $\boldsymbol{\mu}_j$ was more diffuse, but resembles that of $\boldsymbol{\mu}(t)$ (Appendix S5). Parameters in the temporal process model ($\boldsymbol{\beta}$) indicate haul-out use was highest at times near solar noon, during summer months, and at high tide (Appendix S5). Inference concerning animal movement (σ_μ) suggests approximately 95% of at-sea locations were within 6.6 km of a haul-out site. Parameters related to telemetry measurement error are provided in Appendix S5. All inference was based on 50,000 MCMC samples, which required 5 hours of processing time on a computer equipped with a 3.4 GHz Intel Core i7 processor.

Discussion

A fully model-based approach rigorously accommodates multiple sources of uncertainty when estimating the location of central places from satellite telemetry data. Our framework consists of three constituent models: an observation model that accounts for telemetry measurement error and animal movement, a spatial process model for estimating the location of central places, and a temporal process model for quantifying patterns in central place use. Unlike other approaches, our model does not require user-specified distance or time thresholds to identify central places (Anderson and Lindzey 2003), or prior knowledge regarding cluster characteristics (Webb et al. 2008). Model implementation is unified to properly account for uncertainty in parameter estimates.

We demonstrate our model using simulated data examples and an application to harbor seals near Kodiak Island, Alaska. Harbor seals typically exhibit localized movements and regularly return to one or more terrestrial haul-outs between at-sea foraging bouts (Lowry et al. 2001). Our model could also be applied to species that display other behaviors. For example, our model could be used to examine the location of migratory stopover sites or kill sites (Higuchi et al. 2004, Zimmermann et al. 2007. Chevallier et al. 2010); however, the ability to model ephemeral locations requires telemetry data collected at a relatively high temporal frequency.

Observation model

Our observation model consists of a flexible, finite mixture distribution (Eqs. 1 and 3) that accounts for potentially complex telemetry measurement errors like those evident in Argos data (Brost et al. 2015, Buderman et al. 2016). The observation model also accounts for movements away from the central place via an integrated likelihood (Eq. 3; Berger et al. 1999). Because measurement error and animal movement are incorporated into the observation model, we use all telemetry locations to estimate the location of central places, not just those with small magnitude errors or those collected while the individual is at the central place. Furthermore, we use a constrained spatial support for central places (e.g., haul-out sites that only occur along the coastline), and the subsequent discrepancy between the spatial supports of $\mathbf{s}(t)$ and $\boldsymbol{\mu}(t)$, to simultaneously estimate telemetry measurement error (Brost et al. 2015). In applications where central places do not have a constrained support, telemetry error must be known *a priori* or estimated from a secondary data source (e.g., Jonsen et al. 2005, Costa et al. 2010, Douglas et al. 2012).

Process models

The spatial process model consists of a Dirichlet process, a Bayesian nonparametric model that adapts its complexity (e.g., the number of central places) to the observed data. In conjunction with the observation model, the spatial model comprises a Dirichlet process mixture model, a highly flexible framework that includes a large class of distributions (Hjort 2010). As such, the model accommodates multimodal and skewed distributions, like the distribution of central places.

The Dirichlet process allows for potentially infinite clusters as T, the number of observations, approaches ∞; however, the number of occupied components cannot exceed T and is generally much smaller than T. Consequently, a mixture of a finite number of components could be used in practice, which is the strategy we adopt by using a truncation approximation to produce a computationally efficient algorithm for parameter estimation (Ishwaran and James 2001). Other representations of the Dirichlet process, like the Chinese restaurant process, do not rely on truncations for model fitting (Teh et al. 2006).

Our spatial process model could be adapted to include temporal dynamics in the location of central places. For example, seasonal patterns in the location of harbor seal haul-out sites could be incorporated by modeling the central places in a Markovian fashion such that $\boldsymbol{\mu}(t)$ is a function of previous central places. Adjusting our model to differentiate between behaviors would also be necessary if the goal is to examine multiple types of central places in a single dataset (i.e., long-term use of a den site and short-term use of kill sites). One approach to accommodating different behaviors is to formulate the Dirichlet process as a hidden Markov model, a commonly-used method for identifying multiple behavioral states in telemetry data (Patterson et al. 2009, Langrock et al. 2012).

We use a semiparametric regression to model the temporal process of central place use and account for dependence in the behavioral data (Ruppert et al. 2003). Telemetry data are generally not equally spaced in time; thus, serial correlation would be difficult to model using, for example, an autoregressive process. The basis function approach that we implement is a flexible alternative to modeling autocorrelated data (Hefley et al. in revision).

The basis functions, which are continuous in time, also facilitate prediction of animal behavior. For example, animal behavior can be predicted at times associated with telemetry locations when the positional and behavioral data are temporally misaligned (Appendix S6). Our model can even be adapted to estimate animal behavior when ancillary data are not available (Appendix S6). Indeed, prediction is a key advantage of a probabilistic framework like the one we present.

Guidance

The joint analysis of multiple individuals can be achieved by applying our model to several individuals separately, and then combining inference across individuals to obtain population-level parameters with a meta-analysis (e.g., Hartung et al. 2008, Hooten et al. 2016). Alternatively, multiple individuals could be analyzed concurrently using a hierarchical Dirichlet process (Teh et al. 2006, Hjort 2010). A hierarchical approach extends our model by placing individual-specific Dirichlet processes under a common prior (another Dirichlet process), thereby allowing central places to be unique to, or shared amongst, individuals. In either approach, heterogeneity among individuals can be accommodated and explained through the introduction of demographic covariates (e.g., sex and age), and the location of central places could be modeled as a function of environmental covariates to examine site selection.

Bayesian nonparametric models, like the Dirichlet process we use to examine the location of central places, have been adapted to analyze time series data, grouped data, data in a tree, binary data, relational data, and spatial data (Gershman and Blei 2012). This highly flexible framework has been widely used in other fields (Rodriguez and Dunson 2011), although we are aware of few examples from ecology. However, potential ecological applications are numerous and include abundance estimation (Dorazio et al. 2008, Johnson et al. 2013), population genetics (Huelsenbeck and Andolfatto 2007), and disease spread (Verity et al. 2014), among other applications where the goal is to infer latent structure based on empirical data (Morales et al. 2004, Brost and Beier 2012).

Acknowledgements

We thank two anonymous referees for helpful reviews of the manuscript. Our research was funded by the Alaska Department of Fish and Game (ADF&G), through award NA11NMF4390200 from the National Oceanic and Atmospheric Administration Fisheries, Alaska Region to the ADF&G and NSF DMS 1614392. 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

Albert, J. H., and S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association **88**:669–679.

Anderson, C. R., and F. G. Lindzey. 2003. Estimating cougar predation rates from GPS location clusters. The Journal of Wildlife Management **67**:307–316.

Berger, J. O., B. Liseo, and R. L. Wolpert. 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science **14**:1–22.

Blackwell, P. G. 2003. Bayesian inference for Markov processes with diffusion and discrete components. Biometrika **90**:613–627.

Blakesley, J., A. Franklin, and R. Gutiérrez. 1992. Spotted owl roost and nest site selection in northwestern California. The Journal of Wildlife Management **56**:388–392.

Brost, B. M., and P. Beier. 2012. Use of land facets to design linkages for climate change. Ecological Applications **22**:87–103.

Brost, B. M., M. B. Hooten, E. M. Hanks, and R. J. Small. 2015. Animal movement constraints improve resource selection inference in the presence of telemetry error. Ecology **96**:2590–2597.

Buderman, F., M. Hooten, J. Ivan, and T. Shenk. 2016. A functional model for characterizing long-distance movement behaviour. Methods in Ecology and Evolution **7**:264–273.

Chevallier, D., Y. Maho, P. Brossault, F. Baillon, and S. Massemin. 2010. The use of stopover sites by Black Storks (*Ciconia nigra*) migrating between West Europe and West Africa as revealed by satellite telemetry. Journal of Ornithology **152**:1–13.

Costa, D., P. Robinson, J. Arnould, A. Harrison, S. Simmons, J. Hassrick, A. Hoskins, S. Kirkman, H. Oosthuizen, S. Villegas-Amtmann, and D. Crocker. 2010. Accuracy of Argos locations of Pinnipeds at-sea estimated using Fastloc GPS. PLoS ONE **5**:e8677.

Dorazio, R. M., B. Mukherjee, L. Zhang, M. Ghosh, H. L. Jelks, and F. Jordan. 2008. Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. Biometrics **64**:635–644.

Dorazio, R. M., and D. T. Rodríguez. 2012. A Gibbs sampler for Bayesian analysis of site-occupancy data. Methods in Ecology and Evolution **3**:1093–1098.

Douglas, D., R. Weinzierl, S. Davidson, R. Kays, M. Wikelski, and G. Bohrer. 2012. Moderating Argos location errors in animal tracking data. Methods in Ecology and Evolution **3**:999–1007.

Ferguson, T. 1973. A Bayesian analysis of some nonparametric problems. Statistica Sinica **1**:209–230.

Finley, A., S. Banerjee, and A. Gelfand. 2015. spBayes for large univariate and multivariate point-referenced spatio-temporal data models. Journal of Statistical Software **63**:1–28.

Gershman, S. J., and D. M. Blei. 2012. A tutorial on Bayesian nonparametric models. Journal of Mathematical Psychology **56**:1–12.

Hartung, J., G. Knapp, and B. Sinha. 2008. Statistical meta-analysis with applications. John Wiley & Sons, Inc.

Hastie, T., J. H. Friedman, and R. Tibshirani. 2009. The elements of statistical learning data mining, inference, and prediction. Springer, New York.

Hefley, T., K. Broms, B. Brost, F. Buderman, S. Kay, H. Scharf, J. Tipton, P. Williams, and M. Hooten. In revision. The basis function approach to modeling dependent ecological data. Ecology :.

Higuchi, H., J. Pierre, V. Krever, V. Andronov, G. Fujita, K. Ozaki, O. Goroshko, M. Ueta,

S. Smirensky, and N. Mita. 2004. Using a remote technology in conservation: satellite tracking white-naped cranes in Russia and Asia. Conservation Biology **18**:136–147.

Hjort, N. 2010. Bayesian nonparametrics. Cambridge University Press.

Holloran, M., and S. Anderson. 2005. Spatial distribution of Greater Sage-Grouse nests in relatively contiguous sagebrush habitats. The Condor **107**:742–752.

Hooten, M. B., F. E. Buderman, B. M. Brost, E. M. Hanks, and J. S. Ivan. 2016. Hierarchical animal movement models for population-level inference. Environmetrics **27**:322–333.

Huelsenbeck, J. P., and P. Andolfatto. 2007. Inference of population structure under a Dirichlet process model. Genetics **175**:1787–1802.

Ishwaran, H., and L. F. James. 2001. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association **96**:161–173.

Johnson, D. S., P. B. Conn, M. B. Hooten, J. C. Ray, and B. A. Pond. 2012. Spatial occupancy models for large data sets. Ecology **94**:801–808.

Johnson, D. S., R. R. Ream, R. G. Towell, M. T. Williams, and J. D. Leon Guerrero. 2013. Bayesian clustering of animal abundance trends for inference and dimension reduction. Journal of Agricultural, Biological, and Environmental Statistics **18**:299–313.

Jonsen, I. D., J. M. Flemming, and R. A. Myers. 2005. Robust state–space modeling on animal movement data. Ecology **86**:2874–2880.

Knopff, K. H., A. A. Knopff, M. B. Warren, and M. S. Boyce. 2009. Evaluating global positioning system telemetry techniques for estimating cougar predation parameters. Journal of Wildlife Management **73**:586–597.

Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology **93**:2336–2342.

London, J. M., J. M. Ver Hoef, S. J. Jeffries, M. M. Lance, and P. L. Boveng. 2012. Haul-out behavior of harbor seals (*Phoca vitulina*) in Hood Canal, Washington. PLoS One **7**:e38180.

Lowry, L. F., K. J. Frost, J. M. Ver Hoef, and R. A. Delong. 2001. Movements of satellite-tagged subadult and adult harbor seals in Prince William Sound, Alaska. Marine Mammal Science **17**:835–861.

McClintock, B. T., R. King, L. Thomas, J. Matthiopoulos, B. J. McConnell, and J. M. Morales. 2012. A general discrete-time modeling framework for animal movement using multistate random walks. Ecological Monographs **82**:335–349.

McClintock, B. T., J. M. London, M. F. Cameron, and P. L. Boveng. 2014. Modelling animal movement using the Argos satellite telemetry location error ellipse. Methods in Ecology and Evolution **6**:266–277.

Montgomery, R., J. Ver Hoef, and P. Boveng. 2007. Spatial modeling of haul-out site use by harbor seals in Cook Inlet, Alaska. Marine Ecology: Progress Series **341**:257–264.

Morales, J., D. Haydon, J. Frair, K. Holsiner, and J. Fryxell. 2004. Extracting more out of relocation data: Building movement models as mixtures of random walks. Ecology **85**:2436–2445.

Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. Journal of Animal Ecology **78**:1113–1123.

Rodríguez, A., and D. B. Dunson. 2011. Nonparametric Bayesian models through probit stick-breaking processes. Bayesian analysis (Online) **6**:10.1214/11–BA605.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. Semiparametric regression. Cambridge University Press.

Ruprecht, J. S., D. E. Ausband, M. S. Mitchell, E. O. Garton, and P. Zager. 2012.

Homesite attendance based on sex, breeding status, and number of helpers in gray wolf packs. Journal of Mammalogy **93**:1001–1005.

Sethuraman, J. 1994. A constructive definition of Dirichlet priors. Statistica Sinica **4**:639–650.

Teh, Y., M. Jordan, M. Beal, and D. Blei. 2006. Hierarchical Dirichlet processes. Journal of the American Statistical Association **101**:1566–1581.

Tomkiewicz, S. M., M. R. Fuller, J. G. Kie, and K. K. Bates. 2010. Global positioning system and associated technologies in animal behaviour and ecological research. Philosophical Transactions of the Royal Society B **365**:2163–2176.

Verity, R., M. D. Stevenson, D. K. Rossmo, R. A. Nichols, and S. C. Le Comber. 2014. Spatial targeting of infectious disease control: identifying multiple, unknown sources. Methods in Ecology and Evolution **5**:647–655.

Webb, N., M. Hebblewhite, and E. Merrill. 2008. Statistical methods for identifying wolf kill sites using global positioning system locations. The Journal of Wildlife Management **72**:798–807.

Zimmermann, B., P. Wabakken, H. Sand, H. C. Pedersen, and Liberg. 2007. Wolf movement patterns: A key to estimation of kill rate? The Journal of Wildlife Management **71**:1177–1182.

Figure 1. Simulation of 1,000 telemetry locations ($\mathbf{s}(t)$) arising from three central places ($\boldsymbol{\mu}_j$). The point symbology associates telemetry locations (black and gray numerals; most are smaller gray numerals to reduce clutter) to their corresponding central places (white, numbered circles). For example, a telemetry location labeled "1" is associated with the central place labeled "1." The spatial support of central places ($\widetilde{\mathcal{S}}$) exists at the intersection of the blue and gray polygons (black line). The posterior distribution of $\boldsymbol{\mu}(t)$ (red gradient) in the vicinity of the central places is shown in the bottom panels; brighter red corresponds to higher posterior probability. Inference concerning the location of central place "3," which was associated with 608 telemetry locations, is most certain. Inference concerning central places "1" and "2," which were associated with fewer telemetry locations (approximately 200 locations each), is more diffuse. All inference was based on 20,000 MCMC samples after convergence. Note that 326 simulated telemetry locations are beyond the extent of this map, occurring up to 880 km away.

Figure 2. Telemetry locations (top panel) of a subadult female harbor seal monitored from 09 OCT 1995 to 04 JUN 1996 in Ugak Bay (57.42982°N, -152.5715°W) on the southern coast of Kodiak Island, Alaska, USA. Point symbology reflects whether the individual was hauled-out (black points) or at-sea (black crosses) at the time a telemetry location was recorded. Telemetry locations were collected on average every 5.7 h (range: $0.0 - 54.8$ h) using an Argos satellite telemetry device. The animal's position was measured on 1,004 occasions, with $\approx 72\%$ of locations coming from the three least accurate Argos location classes. Approximately 40% of locations were collected while the individual was at a haul-out site ($y(t) = 1$). The spatial support of haul-out sites ($\widetilde{\mathcal{S}}$) exists along the coastline (black line) at the intersection of the blue (water) and gray (land) polygons. The insets show three regions where the posterior probability of $\boldsymbol{\mu}(t)$ (red gradient) is most concentrated (bottom panels). Brighter red corresponds to higher posterior probability. All

inference was based on 50,000 MCMC samples after convergence. Note that 190 telemetry locations are beyond the extent of this map, occurring up to 1,100 km away from Ugak Bay.

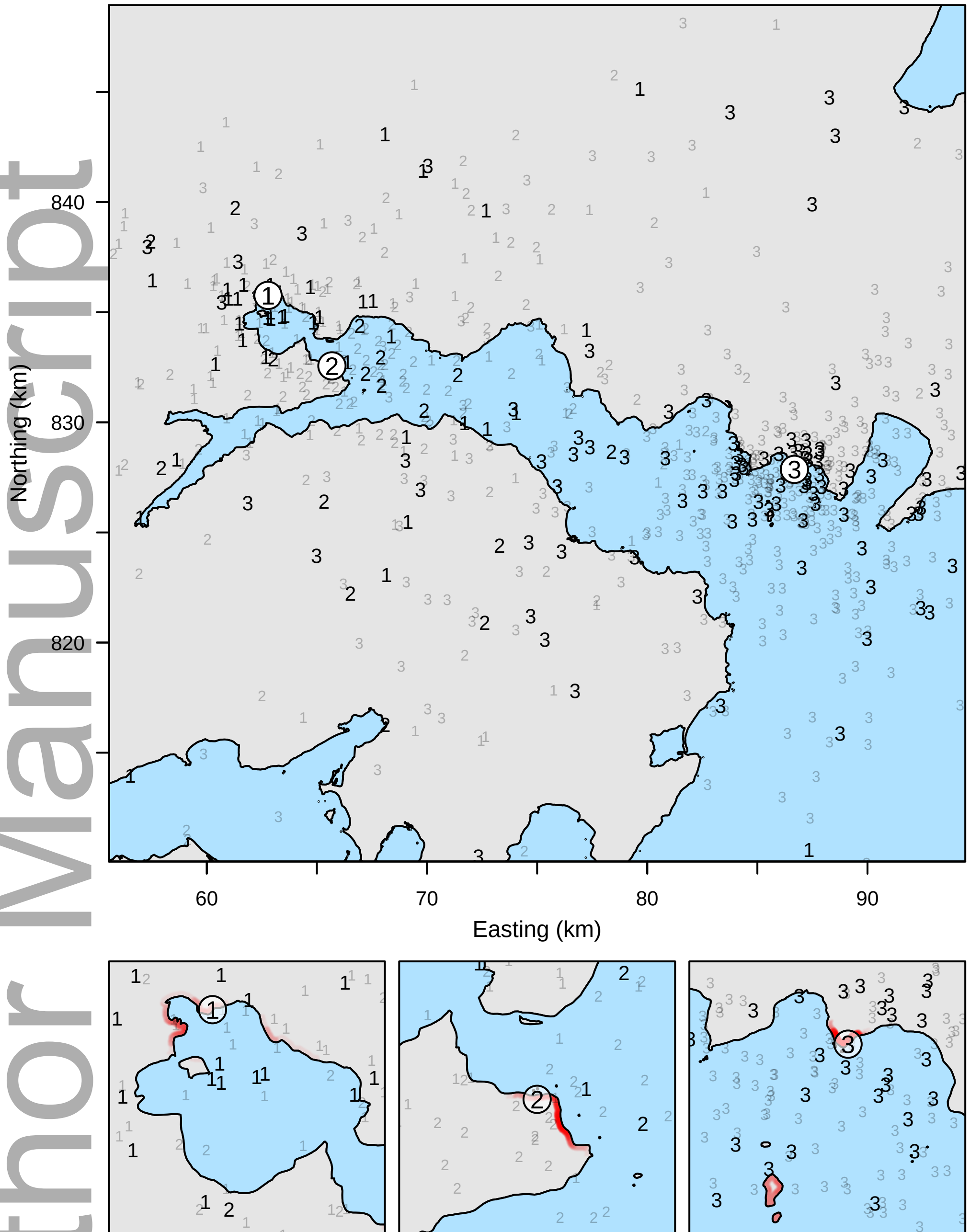

Figure 1:

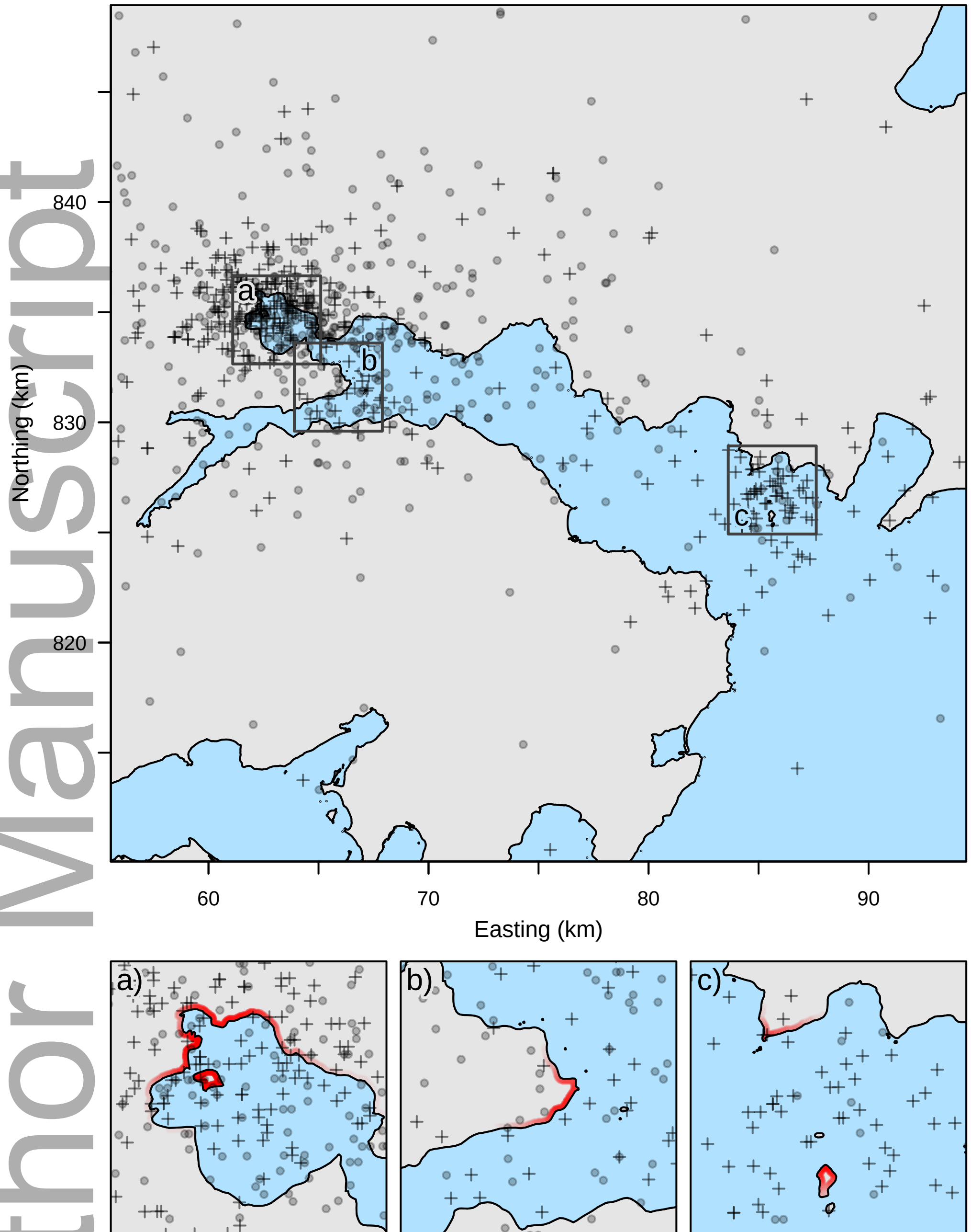

Figure 2: