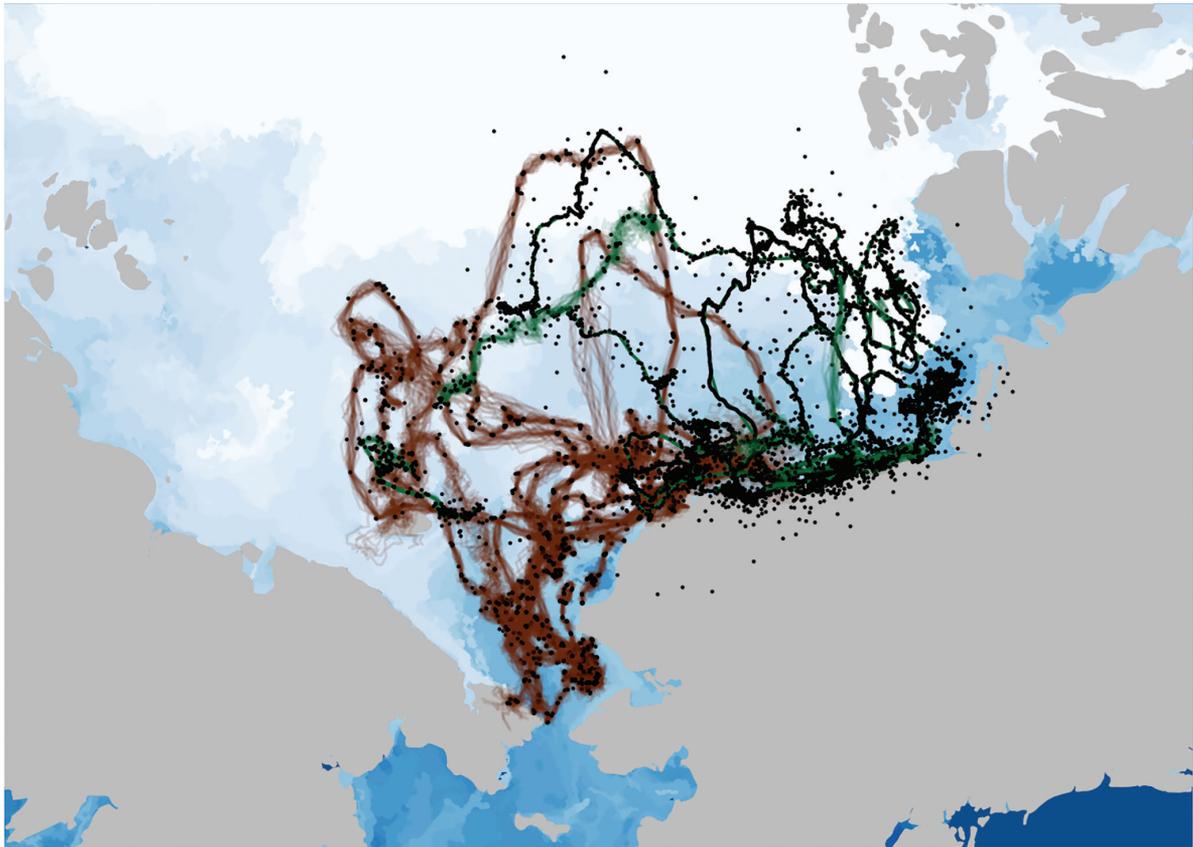

Statistical Analysis of Animal Movement: Understanding Behavior Through Hierarchical Parametric Models



Henry Scharf

Introduction

What are movement data? The movements of people, animals, vehicles, planets, pollen suspended in water, and other objects are fascinating to observe and study because

Henry Scharf is an assistant professor of statistics and data science at San Diego State University. His email address is hscharf@sdsu.edu.

Communicated by Notices Associate Editor Richard Levine.

*For permission to reprint this article, please contact:
reprint-permission@ams.org.*

DOI: <https://doi.org/10.1090/noti2293>

they represent endlessly complex and dynamic relationships in the natural world. The study of motion and trajectories through space and time are relevant for virtually every scientific discipline, especially in ecology where observations of animal movement represent one of the primary opportunities for learning about species' relationships with their environment and each other. With the advent of portable, remotely-sensed devices that can transmit time-stamped measurements of position to distant receivers, it has been possible to track animals' movements in remote locations with increasing frequency, duration,

and precision. The decisions animals make about where and when to move are motivated by many important social, physiological, and external drivers that can have important implications for our understanding of different ecologies. For example, researchers may be interested in learning how seals, often in the open ocean for several days or weeks at a time during foraging excursions, divide their time between searching and feeding behaviors [SHJ17, MB19]. Or, scientists may seek to understand what parts of the landscape facilitate/hinder the movement of mountain lions utilizing kill sites in varied mountain terrain [HHA15, HSM19]. Further, insight into behavior gleaned from movement data can inform national and international policies for responsible management and conservation.

Movement data have historically been gathered in several different ways, but a defining characteristic has been telemetry. That is, observations of trajectories are generally made using remote-sensing technology, wherein a transmitter is affixed to a moving individual, and a receiver records the location of the transmitter at some combination of predefined and opportunistic times. Some commonly used techniques include very/ultra high frequency (V/UHF) radio telemetry, which requires that the receiver and transmitter be connected by line-of-sight, and satellite-based telemetry, in which transmitters can be tracked through hard-to-access parts of the globe such as polar regions and the open ocean.

Advances in battery life and reductions in the size and weight of transmitting devices over the past several decades have brought about a “golden age of animal tracking science” [KCJW15]; however, deploying these devices in the field on wild animals remains a costly and temporarily intrusive process reliant on careful study designs to produce informative data. Moreover, it remains a fundamental challenge to transform raw observations of animals’ positions, usually contaminated with non-ignorable measurement error, into real scientific learning [Nat08]. Thus, there has been a corresponding surge of interest in developing quantitative methods to analyze the current influx of rich movement data. The purpose of this paper is to introduce readers to some of the most common approaches used in the statistical analysis of movement data. Unfortunately, a complete survey of all existing methodologies is not feasible, but several included references offer excellent opportunities for further reading. While the applications that have motivated the research discussed herein have come primarily from ecology, the statistical models are quite flexible and many fundamental mathematical structures are readily adaptable to the study of trajectories arising in other fields of study.

Why statistical models for movement? Several non-statistical methods exist for interpolating observed

locations to produce an estimate of a continuous-time trajectory, so a natural question to ask is what is gained by utilizing statistical models for trajectories over simpler, quicker computational tools. There are three primary benefits.

First, statistical approaches are fundamentally probabilistic and so provide a natural way to quantify uncertainty about inferences made from data.

Second, estimating a trajectory is almost always an intermediate goal. Ultimately, the scientifically interesting characteristics of movement are not the path itself, but what the particular trajectory implies about the interactions between an individual and its environment. For autonomous objects like animals, these interactions are driven by biological and ecological mechanisms that could, for example, be important to understand for the purposes of management and conservation. Statistical approaches offer methods for formally coupling these mechanisms to observations through the use of parametric models in which the values of the parameters inform scientific understanding of animal behavior.

Finally, the statistical models used in the analysis of animal movement are generative, in the sense that it is possible to sample from a probability distribution for the data conditioned on parameter values. Generative models provide a means for checking that relevant mathematical assumptions made in the analysis are compatible with the observed data. In short, it is possible, at least in principle, to assess the “fit” of a model and validate the theory underpinning the mechanisms of interest. In practice, many commonly-used statistical validation methods such as k -fold cross validation are impractical for highly parameterized hierarchical models for movement, but even when gold standard tools for model checking are not implementable, weaker checks can still be investigated.

This paper focuses primarily on the second advantage, connecting observations of a trajectory to scientific learning, and provides some examples of how useful parametric models have been, and continue to be, developed for the study of animal movement. For a more thorough treatment of animal movement modeling from a statistical perspective, see [HJMM17] and references therein.

Parametric Models

In most cases, the goals of researchers studying animal movement include inferring important biological features of behavior for individuals and populations. Statistical approaches provide one way to achieve these goals through the specification of parametric models. A parametric model is a collection, or family, of probability distributions defined by a finite number of parameters. Simple examples of parametric models are location-scale families, such as the family of univariate Gaussian

distributions with unknown mean (location) and variance (scale). This family serves as a parametric model when, for example, it is assumed that a collection of observed data, $\mathbf{y} = (y_1, \dots, y_m)$, are independently and identically distributed according to a Gaussian distribution with unknown mean, $\mu \in \mathbb{R}$, and variance, $\sigma^2 \in \mathbb{R}^+$. Conditioned on this assumption, the data are then used to produce estimates of μ and σ^2 with accompanying measures of uncertainty.

Useful parametric models pose the following qualities: (i) the parameters represent interpretable quantities such that specific values have meaningful implications for the application, (ii) they are flexible enough that, at least for certain combinations of parameter values, the observed data could conceivably represent a realization from the corresponding probability distribution (i.e., the model has “good fit”), and (iii) the family of distributions is sufficiently constrained that the observed data contain information about which collections of parameters are likely to have generated the data (i.e., the parameters are “identifiable”). It is the role of the analyst to specify and “fit” a useful statistical model to data, and then to interpret the estimated parameters. Typically, the process of identifying and fitting a useful parametric model is iterative, and the analyst will need to make several decisions that balance natural tensions between (i), (ii), and (iii) to ensure they are all satisfied.

Likelihood-based inference. Parameters for statistical models can be estimated from data using a variety of techniques. One large class of estimation, or model fitting, techniques is based on the *likelihood* function of the parameters, θ , given the observed data, \mathbf{y} , defined as

$$L(\theta|\mathbf{y}) = p(\mathbf{y}|\theta), \quad (1)$$

where $p(\mathbf{y}|\theta)$ denotes the probability density function for the data evaluated at \mathbf{y} , conditioned on parameter values θ . The functional form of $p(\mathbf{y}|\theta)$ is specified by the analyst as part of the model-building process.

In one paradigm, the analyst assumes that a fixed but unknown set of parameter values, in conjunction with the specified model, generated the data. Estimates of the parameters are obtained by maximizing the likelihood function over the parameter space as $\hat{\theta}^{ML} = \arg \max_{\theta} L(\theta|\mathbf{y})$.

Another perspective regards the parameters as random and specifies a probability distribution, $p(\theta)$, that represents the subjective *prior* beliefs about θ without considering the data [GCS⁺14]. Of interest in this perspective is the *posterior* distribution of θ after considering the data, $p(\theta|\mathbf{y})$, which represents the updated beliefs of the analyst after being confronted with the observed data. The posterior distribution can be computed using the prior

distribution and likelihood function via Bayes’ Theorem as

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \quad (2)$$

$$= \frac{L(\theta|\mathbf{y})p(\theta)}{\int p(\mathbf{y}|\theta)p(\theta)d\theta}, \quad (3)$$

from which Bayesian statistics gets its moniker.

The degree to which the prior distribution influences the posterior compared to the data describes how “informative” the prior is. Intuitively, as the likelihood component of the posterior distribution dominates the product in the numerator of the expression for the posterior distribution, it will have a mode that coincides with $\hat{\theta}^{ML}$ (provided the prior probability density is non-zero there). In some cases, it is possible, though not necessarily desirable, to specify a “non-informative” prior, $p(\theta)$, to align the maximum likelihood estimate and the mode of the posterior distribution. For the vast majority of models, it is also the case that as the number of observations in \mathbf{y} grows toward infinity, the likelihood overwhelms the prior and brings about the same effect.

Parameter estimation for animal movement models has been achieved through both maximum likelihood and Bayesian approaches. While there are meaningful differences in the interpretations and implications of maximum likelihood and Bayesian perspectives, they are united through their foundational reliance on the specification of a likelihood function. In many cases, it is possible to pursue either line of inference for the same core model. The remainder of the paper is organized around similarities and differences among movement models, and, where possible, implementation of each model or class of models is agnostic with respect to a preferred likelihood-based perspective.

Phenomenological Models

Statistical models that describe observed natural phenomena without incorporating scientifically-derived mechanisms, called “phenomenological” models, provide one approach for estimating the true position process with uncertainty from observations. Phenomenological models often have the benefit of requiring only modest assumptions about the data-generating process. As mentioned in the introduction, estimates of a trajectory alone do not generally offer much scientific learning, and therefore stand-alone phenomenological models are of little use to researchers. However, phenomenological models can be used as part of a multistage analytic approach that separates trajectory estimation from behavioral inference in mathematically and computationally convenient ways. In addition, phenomenological models are often a fundamental component in hierarchical model specifications

variable	relationship	definition
$\mathbf{r}(t)$		position at time t
$\varepsilon(t)$		measurement error at t
$\mathbf{s}(t)$	$\mathbf{r}(t) + \varepsilon(t)$	<i>observed</i> position at time t including measurement error
$\mathbf{d}(t)$	$\mathbf{r}(t) - \mathbf{r}(t - \Delta t)$	displacement in position between time $t - \Delta t$ and t
$\mathbf{v}(t)$	$d\mathbf{r}(t)/dt$	velocity at time t
$l(t)$	$\ \mathbf{d}(t)\ _2$	step length between times t and $t - \Delta t$
$\phi(t)$	$\text{sign}(d_1(t)) \arccos\left(\frac{\mathbf{d}(t-\Delta t)'\mathbf{d}(t)}{l(t-\Delta t)l(t)}\right)$	in a Cartesian plane, the turning angle, or change in bearing, between times t and $t - \Delta t$ (note: ' denotes the transpose of a matrix or vector such that $\mathbf{d}(t - \Delta t)'\mathbf{d}(t)$ is the vector inner product)
$\mathbf{x}(\mathbf{r}(t))$		covariates associated with location $\mathbf{r}(t)$
θ		unknown parameters in movement model
$\omega(t)$		latent behavior state (HMM)
$\gamma_{i,j,k}$		probability of transitioning from state j to state k at time i (HMM)

Table 1. Definitions for mathematical notation.

that can be used for joint inference about the true trajectory and the parameters, θ , that enable inference about behavior.

A trajectory through space, $\mathbf{r}(t)$, can be represented mathematically as a multivariate process in time, indexed by t , where each variable corresponds to a dimension in the space of interest. For example, much of animal movement is modeled in the continuous Euclidean plane using Cartesian coordinates, and therefore $\mathbf{r}(t) \in \mathbb{R}^2$ is bivariate with variables corresponding to two orthogonal coordinates (e.g., northing and easting). Polar coordinates and higher dimensional spaces are also used; however, Cartesian coordinates indexed by $c \in \{1, 2\}$ are assumed in the remainder of the paper unless otherwise noted. The time index can be either discrete or continuous. It is most intuitive to think of an animal's true position existing for a continuum of times; however discrete-time indices can also be used, and are often sufficiently good approximations to reality to allow for reliable behavioral inference.

Measurement error. Observations of animal movement data are almost always gathered remotely through the use of transmitters and receivers that introduce measurement error. The scale of measurement error can vary dramatically depending on the measurement device, and may even vary across observations for a single device depending on environmental characteristics (e.g., weather, proximity to Earth's poles, underwater depth, etc.). In some cases, measurement error can safely be ignored, but when it cannot be, a hierarchical model is typically constructed in which the observations are assumed to arise from probability distributions centered on the true position.

Let $\mathbf{x} \sim p(\mathbf{x}|\theta)$ express that a random vector, \mathbf{x} , is distributed according to the probability density p parameterized by θ . The vertical bar notation is used to define a conditional probability density, with variables appearing to

the right of | taken to be fixed. Denote by $N(\mathbf{x}|\mu, \Sigma)$ the probability density function for an n -dimensional multivariate Gaussian (also called normal) random vector, \mathbf{x} , with mean $\mu \in \mathbb{R}^n$ and $n \times n$ covariance matrix Σ . Following this notation, the observed data, $\mathbf{s}(t)$, are assumed to be conditionally independent in time given the true, unobserved location, $\mathbf{r}(t)$, such that

$$\mathbf{s}(t) = \mathbf{r}(t) + \varepsilon(t), \quad (4)$$

$$\varepsilon(t)|\sigma_\varepsilon^2 \sim N(\varepsilon(t)|\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_2). \quad (5)$$

The measurement error process itself, $\varepsilon(t)$, is not typically of interest, but the measurement error variance σ_ε^2 needs to be estimated, unless it is known a priori. It is straightforward to show that for the purposes of estimating σ_ε^2 , the model expressed by (4)–(5) is equivalent to

$$\mathbf{s}(t)|\sigma_\varepsilon^2 \sim N(\mathbf{s}(t)|\mathbf{r}(t), \sigma_\varepsilon^2 \mathbf{I}_2). \quad (6)$$

The likelihood function for the sole parameter σ_ε^2 given observations $\mathbf{s}(t_i), i = 1, \dots, n$, is then

$$L(\sigma_\varepsilon^2|\mathbf{s}(t_1), \dots, \mathbf{s}(t_n)) = \prod_{i=1}^n \prod_{c=1}^2 N(s_c(t_i)|r_c(t_i), \sigma_\varepsilon^2), \quad (7)$$

where the products come from the assumptions of conditional independence in time and across coordinates, c .

The true position process is assumed known in (6) and (7), but of course this would be trivially uninteresting in practice. Thus, a model for $\mathbf{r}(t)$ is required, and its specification represents the central challenge to movement modeling.

Generalized additive models. One approach to modeling $\mathbf{r}(t)$ is to approximate the processes for each coordinate of the true function, $r_c(t)$, as an element in a linear space spanned by functions $h_{j,c}(t)$, $j = 1, \dots, J$, such that $r_c(t) \approx \sum_{j=1}^J a_{j,c} h_{j,c}(t)$. The characteristics of the basis

functions and the dimension of the linear space, J_c , can be specified to ensure the space spans a sufficiently rich collection of functions to provide a good approximation, and that the process obeys reasonable constraints. For example, requiring $h_{j,c}(\cdot)$ to be continuously differentiable implies that $r_c(\cdot)$ will have a continuous first derivative in time representing velocity. Because the particular choice of coordinate axes is arbitrary, a common assumption is that the bases for each spatial coordinate are the same, and so the index c on $h_{j,c}(t)$ and J_c is usually dropped.

If the collection of positions $\mathbf{r}(t)$ at the finite collection of time points t_1, \dots, t_n is vectorized over time for each spatial coordinate (i.e., $\mathbf{r}_c = (r_c(t_1), r_c(t_2), \dots, r_c(t_n))'$), then its approximation in the linear space can be written using matrix notation as $\mathbf{r}_c = \mathbf{H}\mathbf{a}_c$, where \mathbf{H} is a matrix with $h_j(t_i)$ in the i th row and j th column and \mathbf{a}_c is a column-vectorization of the linear coefficients for each spatial coordinate (i.e., $\mathbf{a}_c = (a_{1,c}, a_{2,c}, \dots, a_{J,c})'$).

To fit this model, the coefficients, \mathbf{a}_c , must be estimated along with σ_ε^2 . If the observations, $s_c(t_i)$, are made at times t_1, \dots, t_n with Gaussian-distributed errors, $\varepsilon_c(t)$, centered on the true position with variance σ_ε^2 as in (6), then the statistical model for the observed data is

$$s_c(t) \sim N(s_c(t)|r_c(t) = \sum_{j=1}^J a_{j,c}h_j(t), \sigma_\varepsilon^2), \quad (8)$$

which can be written in a vectorized form as

$$\mathbf{s}_c \sim N(\mathbf{s}_c|\mathbf{H}\mathbf{a}_c, \sigma_\varepsilon^2\mathbf{I}_n). \quad (9)$$

The likelihood function for the full parameter vector $\mathbf{a} = (\mathbf{a}'_1, \mathbf{a}'_2)'$ given observations $\mathbf{s} = (\mathbf{s}'_1, \mathbf{s}'_2)'$ then is

$$L(\mathbf{a}, \sigma_\varepsilon^2|\mathbf{s}) = \prod_{c=1}^2 N(\mathbf{s}_c|\mathbf{H}\mathbf{a}_c, \sigma_\varepsilon^2\mathbf{I}_n). \quad (10)$$

The same independence that allowed the likelihood function to be computed as the product of $2n$ univariate Gaussian densities also applies here. This particular model for movement is a non-linear function of t , but linear in the basis functions. It is therefore an example of multiple linear regression where the predictors are basis functions in time evaluated at a finite number of observation times. Models of this form are known in statistics as generalized additive models (GAMs) [HT86]. Bases in GAMs may be defined for general covariate spaces including temporal indices. When the residuals $\mathbf{s}(t) - \mathbf{r}(t)$ are assumed to be independent and Gaussian-distributed, maximum likelihood estimates are available in a closed form. For most other models for residual variation, numerical optimization techniques are required. More sophisticated GAMs can be specified by allowing for dependence in \mathbf{a} , either through explicit hierarchical construction, or implicitly through penalization (e.g., smoothing splines) [HT86].

In the Bayesian setting, prior distributions are specified for \mathbf{a} and σ_ε^2 . The joint posterior distribution of \mathbf{a} and σ_ε^2 , which has support $\mathbb{R}^{2J} \times \mathbb{R}^+$, does not, in general, have a closed-form expression. Bayesian inference proceeds through one of the most common techniques for implementation called Markov chain Monte Carlo (MCMC). The basic strategy is to build a Markov chain that has as its stationary distribution the joint posterior. The Markov process is then simulated for a sufficiently large number of iterations that can be considered to have converged to this stationary distribution, and realizations from the chain represent draws from the posterior distribution. A large number of posterior draws are accumulated and used to represent the posterior distribution. From this large sample, it is straightforward to summarize marginal distributions of parameters via a central summary statistic and a "credible interval," which is defined to capture a specified proportion of the total probability mass (e.g., 95%) and represents the Bayesian analog to a confidence interval in the frequentist paradigm.

The prior distribution for \mathbf{a} is usually assumed to be Gaussian, primarily because of a convenient mathematical property that facilitates the implementation of MCMC for model fitting. Namely, a Gaussian prior implies that the "full-conditional" distribution of \mathbf{a} given the data and all other unknown parameters, often denoted $p(\mathbf{a}|\cdot)$, is also Gaussian, a joint property of the prior and model known as "conjugacy" in Bayesian statistics. Full-conditional distributions are the central building blocks of the most commonly used MCMC algorithms, which require repeated sampling from these distributions. Conjugate priors have historically been desirable because it is typically simpler to sample from distributions with closed-form density functions than for general distributions. A conjugate prior distribution for σ_ε^2 is the inverse-gamma distribution. In practice, the benefits realized from using conjugate priors are typically modest, thanks largely to improvements in computational resources over the past several decades. The primary consideration when specifying priors should be that the probability distribution align with reasonable a priori beliefs about the model parameters, whatever that might imply about the shape and support of the distribution [GCS⁺14].

An example of GAMs used in a Bayesian analysis of animal movement is [BHIS16]. The authors implemented a two-stage approach to study the dispersal of a reintroduced population of Canada lynx (*Lynx canadensis*) in Colorado. A GAM-based phenomenological model was used to link models for dispersal behavior to the observed locations of two Canada lynx over two years.

Application: Fin whales (part I)

Figure 1 shows the observed locations of a fin whale (*Balaenoptera physalus*) made at irregular intervals, varying

from a few minutes to a few hours, during July 17–July 31, 2015, on a map of the California coast from Los Angeles to San Jose [IWF⁺20]. Points represent the observed locations, \mathbf{s} , the line represents the maximum likelihood estimate of the true position, \mathbf{r} , based on (10), and the ellipses represent pointwise 95% confidence regions at regular one-hour intervals during the two-week study period based on the curvature of the likelihood function near its maximum. The plots in the top right are the marginal positions of the individual in each spatial dimension. The basis functions, $h_j(t)$, are cubic regression splines, which are one of the most commonly used bases in GAMs [Woo20]. Estimates of the continuously indexed trajectory taken by an individual provide limited scientific learning on their own, but can be a useful part of a multistage analysis as shown in Application: Fin whales (part III). Data for all applications are freely available online, and code that reproduces all analyses and figures is available at https://github.com/henryrscharf/AMS_statistical_movement.

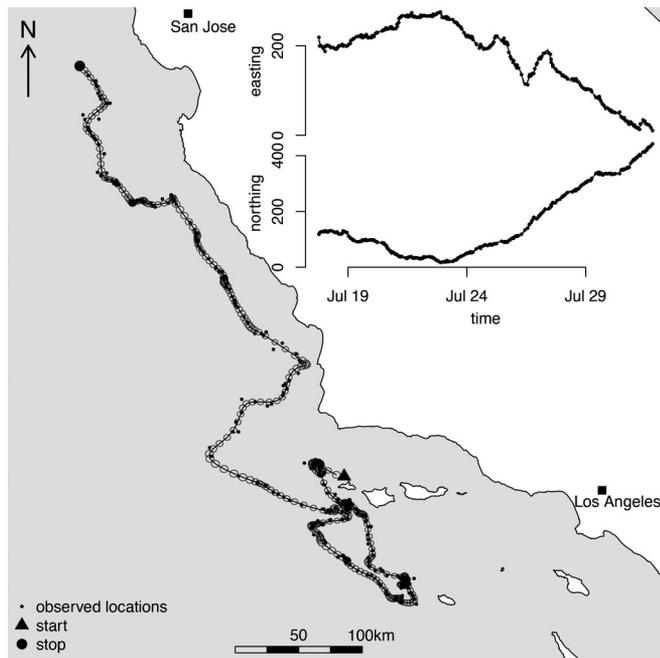


Figure 1. Observed locations (points), estimated path (line), and associated pointwise 95% confidence ellipses for a fin whale during July 18–July 31, 2015. The two plots in the upper right show the marginal observations (points) and estimated location (line) in the horizontal (easting) and vertical (northing) coordinates over time.

State-space models. Specifying a collection of basis functions, and corresponding linear space, represents a global perspective on animal movement in which characteristics of the entire modeled trajectory are defined by the choice of $h_j(t)$. An alternative is to focus on modeling transitions, or differences in position over short time intervals, and let

the global properties arise from specifications of the local behavior.

For continuous-time movement models, a necessary constraint for an animal’s trajectory is that it be continuous in space. An animal should not be allowed to “jump” positive distances instantaneously. The Wiener process, also called Brownian motion, is one of the simplest processes that satisfies the continuity constraint. The standard Wiener process indexed by $t \in [0, \infty)$ is defined such that each infinitesimal displacement is Gaussian-distributed with variance dt , and independent of all other displacements not overlapping in time with the interval $(t - dt, t)$. Let $d\mathbf{r}(t) = \mathbf{r}(t) - \mathbf{r}(t - dt)$ represent the displacement of an individual over an increment of time dt . A simple model for the displacement $d\mathbf{r}(t)$ that ensures continuity can be written as

$$\begin{aligned} d\mathbf{r}(t) &= \sigma_\eta d\mathbf{W}(t), \\ d\mathbf{W}(t) &\sim N(\mathbf{0}, dt\mathbf{I}_2), \end{aligned} \quad (11)$$

where $d\mathbf{W}(t)$ aligns with the conventional notation used in the literature for increments of a standard Wiener process, and σ_η scales the process to the appropriate units for movement.

A Wiener process model can be paired with the same measurement error model in (6) to yield a phenomenological model for imperfectly observed movement parameterized by σ_ε^2 and σ_η . The variable σ_η has an intuitive interpretation as the typical step length, or magnitude of the displacement per unit time. One commonly encountered complication to implementation of a movement model described by (6) and (11) is difficulty simultaneously estimating σ_ε^2 and σ_η . For this reason, it is common in practice to either fix the measurement error variance, σ_ε^2 , or specify an informative prior for it based on a priori knowledge about the telemetry devices used to gather the data.

The model for observations of a Wiener process contaminated by Gaussian error is a special case of a collection of stochastic models called state-space models (SSMs). SSMs have long been used to analyze, interpolate, and predict trajectories observed using devices that introduce non-ignorable measurement error, so it is no surprise that many hierarchical animal movement models for the unobserved true trajectory and observation process fit the form of an SSM [PTW⁺08]. The Wiener process model with Gaussian measurement error is special because the conditional distribution for $\mathbf{r}(t_2)|\mathbf{r}(t_1)$ for any $t_1 < t_2$ is available in closed form, and therefore a hierarchical model for observations $\mathbf{s}(t_1), \dots, \mathbf{s}(t_n)$ can be written

$$\mathbf{s}(t_i) = \mathbf{r}(t_i) + \varepsilon(t_i), \quad \varepsilon(t_i) \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}_2), \quad (12)$$

$$\mathbf{r}(t_i) = \mathbf{r}(t_{i-1}) + \eta(t_i), \quad \eta(t_i) \sim N(\mathbf{0}, \sigma_\eta^2 \Delta(t_i) \mathbf{I}_2), \quad (13)$$

where $\Delta(t_i) = t_i - t_{i-1}$. Though the model for the latent position is indexed continuously in time, (12) and (13)

make it clear that the latent movement process only needs to be considered at the n observation times. Notation is often simplified slightly by using subscripts for variables evaluated at one of the n time points (e.g., $\mathbf{r}(t_i) = \mathbf{r}_i$).

The characteristics of the Wiener process model with Gaussian measurement error model that enable this representation are the Gaussianity of ε_i and η_i , and the linear relationships between \mathbf{s}_i and \mathbf{r}_i and between \mathbf{r}_i and \mathbf{r}_{i-1} . A general class of linear Gaussian SSMs can be represented as

$$\mathbf{s}_i = \mathbf{Z}_i \alpha_i + \varepsilon_i, \quad \varepsilon_i \sim \mathbf{N}(\mathbf{0}, \mathbf{S}_i), \quad (14)$$

$$\alpha_i = \mathbf{T}_i \alpha_{i-1} + \eta_i, \quad \eta_i \sim \mathbf{N}(\mathbf{0}, \mathbf{R}_i), \quad (15)$$

where α_i can represent any general process and is called the “state” of the system at time t , and \mathbf{s}_i is the observation process related to α_i . Equations (14) and (15) are referred to as the observation and state equations, respectively. For the case of animal movement, it is natural to think of \mathbf{s}_i as the observed location. The state vector, α_i , often corresponds to the true location, \mathbf{r}_i , but can also be augmented to include additional state information such as velocity [JLLD08], hence the change in notation. The linearity of this model refers to the linear operators \mathbf{Z}_i and \mathbf{T}_i , and the Gaussianity refers to the distributions of ε_i and η_i . The matrices \mathbf{Z}_i , \mathbf{S}_i , \mathbf{T}_i , \mathbf{R}_i may all depend on unknown parameters θ .

In a typical animal movement application, the primary research objectives are usually to estimate the true trajectory, \mathbf{r} (potentially at times other than the observation times), and model parameters, θ , with uncertainty. In the Bayesian paradigm, this means estimating the posterior distributions $\alpha_{1:n} | \mathbf{s}_{1:n}$ (called the “smoothing” distribution) and $\theta | \mathbf{y}_{1:n}$. In applications of SSMs outside of animal movement (e.g., return series in finance, movements of autonomous vehicles), there is often an interest in predicting the state process one time-step beyond the last observation. The distribution of the state one step into the future, $\alpha_i | \mathbf{s}_{1:i-1}$, is called the “filtering” distribution. The filtering distribution is not commonly of interest in animal movement studies as data are often analyzed after the observation process has been terminated. However, one potential application of the filtering distribution to movement might be for wildlife managers interested in predicting abrupt changes in animal behavior in real time. For example, managers might be interested in online predictions of when female deer or elk give birth so that they can intercept and tag the new fawn for future study.

The Kalman filter [Kal60, DK01] and smoother are methods for efficiently computing the filtering and smoothing distributions of the latent process α conditioned on θ when the SSM is linear and Gaussian. Because the likelihood function of an SSM can be decomposed into the product-filtering distributions, the Kalman filter also

provides efficient calculation of the likelihood, which can be used to estimate θ . Non-linear SSMs have not commonly been used for animal movement, although they are common in other fields. Non-linearity generally necessitates approximate inferential procedures that use, for example, the extended Kalman filter, particle filters [DdFG01], or ensemble Kalman filters.

Application: Fin whales (part II)

While the Wiener process model does satisfy the continuity constraint for the latent position process, in practice, it is unrealistic for animal movement unless used for very coarse temporal resolutions. Realizations of the Wiener process are nowhere differentiable in time, which implies that the instantaneous velocity process, $\mathbf{v}(t) = d\mathbf{r}(t)/dt$, is not well defined. Visually, this lack of analytical smoothness in Wiener process realizations manifests itself as trajectories that are too “rough” to be plausible for the movement of animals.

One way to generate smoother trajectories is to integrate a rough process in time, which explicitly increments the number of continuous derivatives by 1. Equivalently, a process too rough to represent position could be specified for velocity instead. Johnson et al. [JLLD08] developed a widely used phenomenological model for movement by specifying an Ornstein-Uhlenbeck (OU) process for the velocity of an animal. The OU process is similar to the Wiener process, except that it has a “mean reversion” property that tends to keep realizations near a fixed value, γ . The OU process can be defined by the evolution of the process over an infinitesimal increment of time as

$$d\mathbf{v}(t) = \beta(\gamma - \mathbf{v}(t))dt + \sigma_\eta d\mathbf{W}(t), \quad (16)$$

where $\beta \geq 0$ controls the mean reversion behavior. Like the Wiener process, the differential representation of the OU process can be solved to yield the following conditional distribution for any time points $t_{i-1} < t_i$ and $\Delta_i = t_i - t_{i-1}$:

$$\mathbf{v}_i | \mathbf{v}_{i-1} = \gamma + e^{-\beta \Delta_i} (\mathbf{v}_{i-1} - \gamma) + \eta_i, \quad (17)$$

$$\eta_i \sim \mathbf{N}(\mathbf{0}, \sigma_\eta^2 (1 - e^{-2\beta \Delta_i}) / 2\beta \mathbf{I}_2). \quad (18)$$

The Wiener process arises as a special case when $\beta = \gamma = 0$.

A challenge that arises with specifying a process for velocity is that the process model and the measurement model are incompatible, as the measurements are of position not velocity. In [JLLD08], the authors showed how this challenge could be met by defining the state of the system to include both the position and velocity. For $\alpha_i = (\mathbf{r}_i, \mathbf{v}_i)'$, the authors derive conditional distributions for $\alpha_i | \alpha_{i-1}$, which are also Gaussian and linear in α_{i-1} . Thus, by specifying $\mathbf{Z}_i = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, observations can be matched to positions implied by the OU velocity process.

par.	pt. est. (95% CI)
σ_3	0.36 (0.001, 100)
σ_2	1.09 (0.78, 1.53)
σ_1	1.20 (0.99, 1.44)
σ_0	1.75 (1.52, 2.01)
σ_A	0.98 (0.83, 1.16)
σ_B	1.39 (1.32, 1.46)
τ	2.67 (2.39, 2.99)
β	0.31 (0.21, 0.46)

Table 2. Maximum likelihood estimates (and confidence intervals) for model parameters. Standard deviations, rather than variances, are reported.

The gray curves in Figure 2 show 32 realizations from the movement process from [JLLD08] fit to the fin whale data using maximum likelihood via the Kalman filter. The blue and yellow curve represents the pointwise mean of the 32 trajectories, whose colors are discussed in the next section. The observation error variance, σ_{η}^2 , was generalized slightly to account for the Argos system used to measure position. Argos data are recorded using polar-orbiting satellites with associated uncertainty that can vary dramatically across the observations. Locations are reported along with one of six error classes (from least to most severe: 3, 2, 1, 0, A, B). A separate observation error variance was estimated for each of the six error classes. Maximum likelihood point estimates and 95% confidence intervals for standard deviations are given in Table 2. The estimates generally follow Argos' severity ordering, with the exception of classes 0 and A.

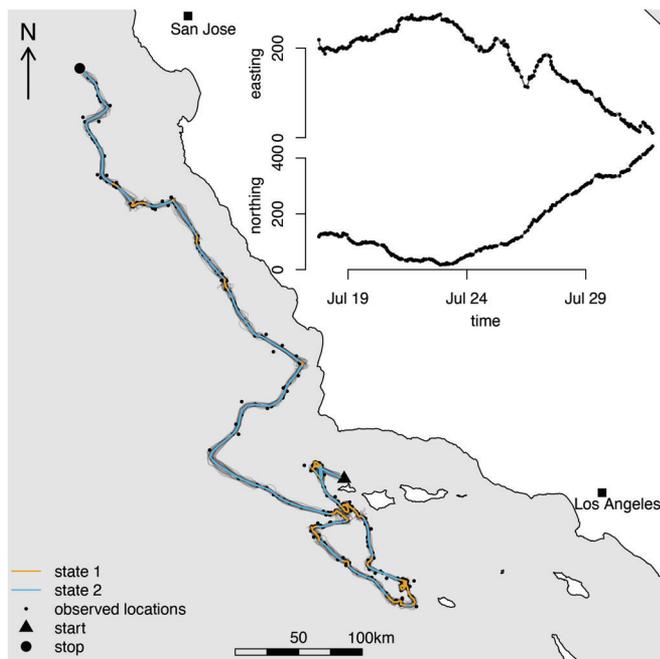


Figure 2. This is the fit using an OU model for velocity.

Visualization of trajectory uncertainty. A persistent challenge in the analysis of movement data is the display of joint uncertainty for the true trajectory. A standard approach for summarizing position uncertainty is to provide confidence/credible regions around a point estimate for the true location of an individual across a grid of times, where each region is derived marginally, or pointwise, in time. Figure 1 shows an example of this approach in which each region is a 95% confidence ellipse produced at one-hour intervals.

Providing pointwise uncertainty quantification is most useful when there is interest in estimating locations at well-separated time points, ideally identified prior to the analysis. Pointwise uncertainty summaries are less useful for portions of a trajectory that occur over a substantial duration of time because they fail to capture dependence between temporally proximate positions. Ignoring temporal dependence can lead to incomplete or incorrect intuitions about what sorts of trajectories are probable because it masks higher-order aspects of the movement process. For instance, it might seem intuitively reasonable that a path constructed by linearly interpolating points chosen at random from each confidence region in Figure 1 would be among the 95% of paths most compatible with the data. However, such a construction ignores dependence in time and might well generate a path whose implied velocity process was virtually impossible. A better way to graphically represent uncertainty for trajectories is to plot several realizations of plausible trajectories, as in Figure 2, rather than a single estimate and pointwise uncertainty.

Mechanistic Models

Phenomenological models excel at explaining *how* an individual moves and are particularly well-suited to predictive tasks such as interpolation. They are also useful for multistaged approaches to inference, and one such example is explored below. The fundamental limitation of phenomenological models is that they lack any mechanistic core that could help explain *why* an individual moves the way it does. Mechanistic models, in contrast, can provide specific, scientifically valuable learning about why individuals use space the way they do. However, they necessarily require stronger assumptions about the probabilistic underpinnings of statistical movement models. These assumptions are typically expressed as behavioral rules that explain some, but never all, of the biological processes driving movement.

HMMs: Models for multiple behavior states. One common example of a simplified, but nevertheless valuable, and frequently-used animal movement model describes trajectories conditioned on a sequence of latent behavior states that manifest themselves in observable characteristics of movement. The sequence of behaviors is modeled

using a discrete-time Markov process, and the movement process is assumed to be conditionally independent in time given the latent behaviors. Thus, the approach falls into the broad class of statistical models known as hidden Markov models (HMMs), for which robust theory and implementation techniques have been thoroughly developed [ZML16].

In the context of movement, the position process, $\mathbf{r}(t)$, is typically reparameterized into step lengths, or the magnitude of the displacement vector between two consecutive times, and turning angles, or the change in bearing, as it is typically easier to connect these features of a trajectory to particular behaviors than the raw positions. For example, three behavior states commonly used to summarize movement are “encamped,” which generally refers to meandering movement with small step lengths and large variability in turning angle, “transit,” which is characterized by large step lengths and small turning angles that produce directed movement, and “resting,” for which step lengths are 0. In a Bayesian paradigm, it is possible to specify priors for the parameters governing the distributions of step lengths and turning angles that explicitly imbue state labels with meaning reflecting a priori beliefs about behavior. In practice, it is much more common to prescribe weakly informative priors or adopt a frequentist paradigm, and assign behavior state labels post hoc.

In conjunction with an initial location and bearing, a sequence of step lengths, $l(t_i)$, $i = 1, \dots, n$, and turning angles, $\phi(t_i)$ (see Table 1), uniquely defines a trajectory at the n time points. The distributions for the step length and turning angle between two consecutive times are specified conditionally on the unobserved behavior state, $\omega(t_i)$, which can take a finite number of possible values corresponding to different types of activity. The Markov behavior state process transitions from state j to state k at time t_i with probability $\gamma_{i,j,k}$.

Adopting the subscript index notation, the HMM for step lengths and turning angles conditioned on behavior ω_i can be described mathematically as

$$l_i | \omega_i = j \sim \mathcal{F}_l(l_i | \theta_i(j)), \quad (19)$$

$$\phi_i | \omega_i = j \sim \mathcal{F}_\phi(\phi_i | \theta_i(j)), \quad (20)$$

$$\Pr(\omega_i = k | \omega_{i-1} = j) = \gamma_{i,j,k}, \quad (21)$$

where $\mathcal{F}_l(l_i | \theta_i(j))$ and $\mathcal{F}_\phi(\phi_i | \theta_i(j))$ are the conditional distributions of step length and turning angle given the individual is in behavior state $j \in \{1, \dots, J\}$.

The transition probabilities can be allowed to vary in time by introducing another layer to the hierarchical model. Often, it is reasonable to assume that features of the environment near an individual affect the probability of transitioning from one behavior state to another. One model for transition probabilities as functions of environmental covariates, $\mathbf{x}(\mathbf{r}_i)$, associated with the spatial

location \mathbf{r}_i is

$$\gamma_{j,k}(\mathbf{x}) = \frac{\exp(\mathbf{x}'\boldsymbol{\beta}_{j,k})}{\sum_k \exp(\mathbf{x}'\boldsymbol{\beta}_{j,k})}, \quad (22)$$

which implies $\gamma_{i,j,k} = \gamma_{j,k}(\mathbf{x}(\mathbf{r}_i))$. Positive elements in the vector $\boldsymbol{\beta}_{j,k}$ indicate those covariates that increase the probability of transitioning from state j to state k , and negative elements indicate covariates that decrease transition probabilities.

For observations made at regular intervals with negligible measurement error, the sequence of step lengths and turning angles can be calculated from the observations and model fitting is straightforward and computationally inexpensive. However, the introduction of substantial measurement error complicates inference, as the simple additive noise used in previously described models cannot be applied directly to step lengths and turning angles. The time index for the HMM for movement given by (19)–(21) must necessarily be regular, as irregular time steps would induce an inconsistent interpretation of the transition probabilities. However, a discrete-time setting does not accommodate irregular observations of position. Both of these limitations of HMMs can be addressed by adopting a two-stage approximate inferential approach, as outlined in the following section. Alternatively, recently developed continuous-time movement models that incorporate multiple behavior states have the potential to merge useful aspects of SSMs and HMMs [MB19].

Multiple imputation

In some cases, observations of animal movement are made at nearly regular intervals with sufficiently small measurement error that a discrete-time HMM can be applied directly to the data. When this is not true, it is a challenge to define a data-level model in the hierarchical structure, $p(\mathbf{s}|\mathbf{r})$, that can reconcile a discrete-time HMM with real data. One approach that has been used with some success is to employ a two-stage procedure in which a continuous-time phenomenological model helps bridge the gap between process and data. In the first stage, a phenomenological model is fit to the data, which provides a way to obtain K realizations from $\mathbf{r}|\mathbf{s}$ on a regular grid of times. These realizations are then treated as “data” in the second stage, where the desired mechanistic model can be used to infer the values of model parameters, $\theta^{(k)}$, conditioned on each $\mathbf{r}^{(k)}$, and aggregated across K . Relying on $K > 1$ realizations, $\mathbf{r}^{(k)}$, propagates uncertainty about the true position process through to the second stage of inference.

The approach is motivated by a decades-old method for missing data called multiple imputation, which was originally developed for Bayesian analysis [Rub96]. For animal movement, the role of “missing data” is played by the true position process \mathbf{r} . The goal is to estimate the posterior distribution of θ given the observations \mathbf{s} , which would

normally be achieved by specifying distributions for $p(\mathbf{s}|\mathbf{r})$ and $p(\mathbf{r}|\theta)$ in a hierarchical model, and using an MCMC algorithm to approximate the integral required for the likelihood

$$p(\theta|\mathbf{s}) = \int p(\theta|\mathbf{r})p(\mathbf{r}|\mathbf{s})d\mathbf{r}. \quad (23)$$

This equation makes plain the misalignment between the data and the latent position process. To admit a density for $\mathbf{s}|\mathbf{r}$, \mathbf{r} must be irregularly spaced in time, but the index must be regular when the model for $\mathbf{r}|\theta$ is an HMM. If it were possible to sample from $p(\mathbf{r}|\mathbf{s})$ on the discrete-time grid, then Monte Carlo estimates of the integral in (23) could be obtained by evaluating $\theta|\mathbf{r}^{(k)}$ for $k = 1, \dots, K$ realizations by

$$p(\theta|\mathbf{s}) = K^{-1} \sum_{k=1}^K p(\theta|\mathbf{r} = \mathbf{r}^{(k)}). \quad (24)$$

To address the mismatch, an alternative “approximate imputation distribution” (AID), $p(\tilde{\mathbf{r}}|\mathbf{s}, \psi)$, is substituted for $p(\mathbf{r}|\mathbf{s})$. To accommodate irregular observations, AIDs are generally continuously indexed in time. Draws from the AID, $\tilde{\mathbf{r}}^{(k)}$, are then evaluated only at the discrete-time index required by the HMM and substituted in (24).

AIDs come from a family of distributions parameterized by variables ψ estimated from the data. The process of first estimating ψ and then using realizations from the AID to estimate θ constitutes the two stages in the approximate inferential procedure. A good AID is able to generate nearly all realistic trajectories passing sufficiently close to the observations to be consistent with a realistic amount of measurement error. The most commonly used AID is based on the integrated OU process with Gaussian measurement error of [JLLD08], in which case $\psi = (\beta, \tau^2, \sigma_\epsilon^2)$. The integrated OU process model is fit to the data by maximizing the likelihood computed using the Kalman filter. The smoothing distribution for position conditioned on either the maximum likelihood point estimates of ψ , or integrated over a range of probable values for ψ , then serves as the AID.

The use of an AID has the following heuristic justification. In the limit that the observation times align with the grid implied by $p(\mathbf{r}|\theta)$, and measurement error diminishes to 0, the probability density $p(\mathbf{r}|\mathbf{s})$ becomes a point-mass at $\mathbf{r}_i = \mathbf{s}_i, i = 1, \dots, n$. For small deviations in observation times from the regular grid and modest measurement error, $p(\mathbf{r}|\mathbf{s})$ remains highly concentrated around \mathbf{s} . For any continuous-time AID that is consistent with the true severity of measurement error, $p(\tilde{\mathbf{r}}|\mathbf{s}, \psi)$ will also be concentrated on trajectories that pass close to \mathbf{s} . This will be true for a wide variety of dependence structures in the joint distribution for $\tilde{\mathbf{r}}$. For example, both a linear interpolation and a functional approximation using infinitely

differentiable basis functions could perfectly interpolate a finite sequence of observed locations, even though they imply very different behavior for the continuous velocity process $\tilde{\mathbf{v}}(t) = d\tilde{\mathbf{r}}(t)/dt$. Because those differences are not relevant for the discrete-time model $p(\mathbf{r}|\theta)$, the approximation error introduced by the AID is negligible for nearly regular observations with low measurement error.

The validity of multiple imputation with an AID deteriorates as observations become less regular, and/or measurement error increases. Approximation errors in the two-stage inferential procedure will manifest themselves to a greater degree for some AIDs more than others, and will have a varied impact on each element θ . For example, commonly used AIDs implicitly involve a single behavior state, which could lead to biased inference when paired with a multistate second stage model. In the few dedicated investigations into the performance of multiple imputation in animal movement models, the integrated OU process of [JLLD08] has generally been seen to perform the best. The elements of θ least sensitive to the approximations introduced by multiple imputation tend to be parameters that depend most directly on the position process, while parameters related to velocity or acceleration can be highly sensitive to the choice of AID, and should generally be interpreted more cautiously [SHJ17]. Finally, it should be noted that the two stages used in multiple imputation are not formally generative in that they do not explicitly define a conditional distribution for the data given the second-stage parameters, $p(\mathbf{s}|\theta)$. Recently, promising new approaches to inferring multistate behavior from irregular, imperfect observations while avoiding multiple imputation entirely have been proposed [MB19], offering a potential alternative to multiple imputation used in conjunction with HMMs.

Application: Fin whales (part III)

The realizations from the integrated OU process shown in Figure 2 were used to approximately fit a HMM to the fin whale data. Two behavior states, $j \in \{1, 2\}$, were assumed, and transition probabilities were modeled as functions of ocean depth [Cen] according to (22). The distributions for step lengths and turning angles were specified as Gamma and von Mises distributions, respectively, with different parameters for each behavior state that were assumed to be constant in time. The prior distributions for both $\theta(j = 1)$ and $\theta(j = 2)$ were specified such that the mean a priori step length was 2 kilometers per one-hour time interval with a standard deviation of 7 kilometers, and turning angles had uniform probability densities on $(-\pi, \pi]$.

Figure 3 shows summaries of the empirical densities of step lengths and turning angles (gray histograms), posterior densities conditioned on the two behavior states (orange and blue curves), and a weighted average of the conditional distributions based on the proportion of time

the individual was estimated to have spent in each behavior state (dashed black curves). The two inferred behaviors appear to correspond to encamped (orange) and transitory (blue) states.

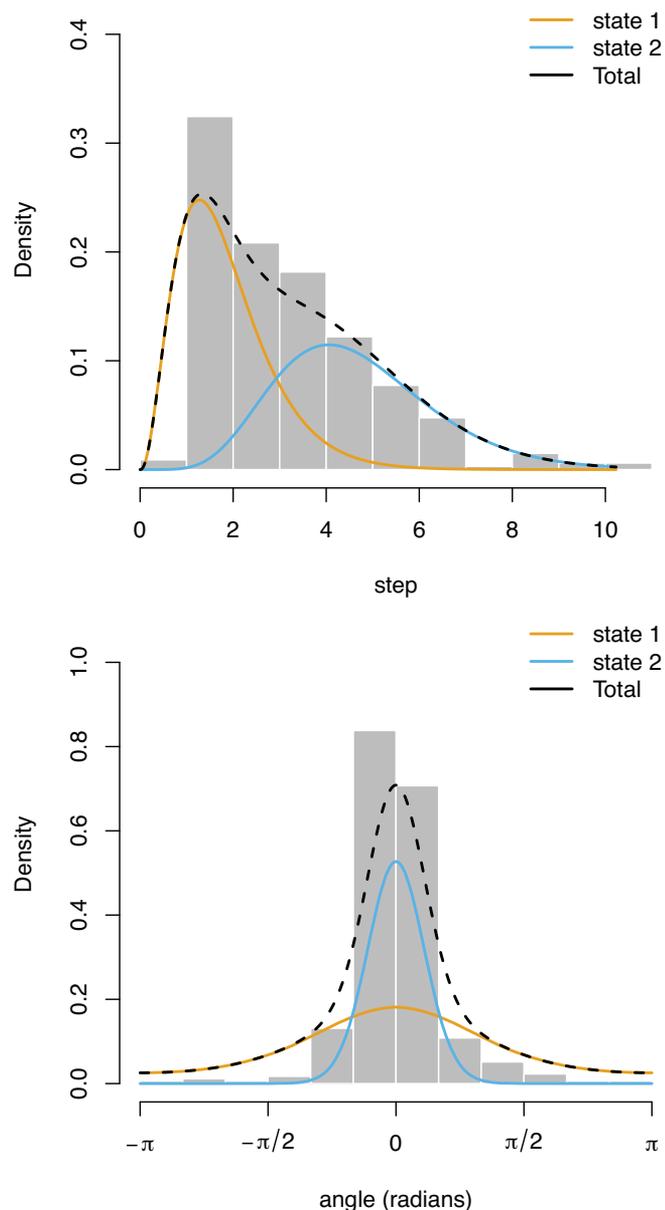


Figure 3. Empirical distribution of step lengths (top; gray histogram) and turning angles (bottom) with posterior distributions for each behavior state (colored curves) overlaid. The dashed black line represents a weighted average of the posteriors based on the proportion of time spent in each state in the Viterbi-decoded state sequence.

Transition probabilities were modeled as functions of ocean depth (measured in kilometers) according to (22). Figure 4 shows the posterior mean of $\gamma_{j,k}(x_{\text{elev}})$ with pointwise 95% confidence intervals. The whale appears to

change behaviors infrequently (as evidenced by high values on diagonal plots), with a moderate increase in the tendency to transition into, and occupy, an encamped state as ocean depth decreases (i.e., shallower water).

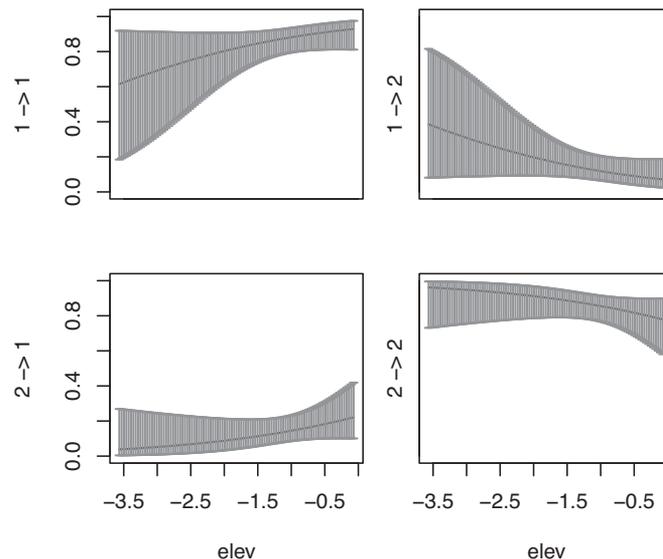


Figure 4. Transition probabilities for the two behavior states (see Figure 3) as functions of ocean depth, shown with 95% confidence intervals.

Once estimates of the parameters in each of the two state-dependent distributions for step length and turning angle are made, it is possible to obtain the highest probability sequence of states for the HMM using the Viterbi algorithm [ZML16]. The most probable sequence of behaviors is not necessarily the same as the pointwise mode of the behavior process $\omega = (\omega_1, \dots, \omega_n)$. The blue and yellow trajectory in Figure 2 represents the pointwise mean of the 32 realizations colored according to the Viterbi-decoded state sequence using the modes of the posterior distributions for $\theta(1)$ and $\theta(2)$.

HMMs for multiple trajectories. HMMs have also been developed for the joint movement of interacting individuals appropriate for describing social animals such as reindeer (*Rangifer tarandus*) and killer whales (*Orcinus orca*). Migrating reindeer move in herds driven by seasonal changes in vegetation, typically staying close together to remain less vulnerable to predators. Occasionally, individuals will move independently of the herd in response to competition for scarce food sources. To capture these two behavior states, an HMM was developed for regular, error-free observations of 11 individual reindeer in Sweden [LHB⁺14] occupying one of two possible behaviors, summarized as “group” and “individual” movement. Behavior was individual-specific, in the sense that it was possible for some portions of the herd to act as individuals while the rest moved as a group. One way to represent the collection

of possible behaviors is as dynamic graphical networks with 11 vertices representing the individuals in the population, and edges representing connections between socially-interacting individuals. Dynamic networks for this behavior model are composed of a single completely connected subgraph, and a complementary set of unconnected vertices. One conclusion drawn from the analysis of reindeer movement using this approach was that individuals primarily occupied a group behavior state, sometimes for several consecutive days.

More flexible graphical structures have been used in applications to the movements of killer whales [SHF⁺16]. The movement of seven whales from three different subtypes was modeled using an HMM conditioned on a binary, undirected social network describing the instantaneous pairwise connections between individuals in the group. Given the dynamic latent social network, changes in individuals' positions were modeled as the sum of attraction and alignment with social neighbors. The strength of the effect of the social network on attraction and alignment was parameterized and estimated from the data in conjunction with the latent dynamic network. A Markov model for the latent social state of the population ensured that the social dynamics varied on a reasonable time scale.

One of the primary goals that inspired the HMM for multiple individuals in [SHF⁺16] was to infer a dynamic social network from movement data, as direct observations of social interactions can be expensive or impossible to record for species in remote locations. It was shown both in simulation and in an application to killer whales that the inferred social network based on the HMM was consistent with scientific knowledge, while alternative non-statistical approaches to network estimation based on relative proximity were not.

Continuous-time discrete space. Another class of mechanistic models provides useful links between movement and environmental covariates by adopting a continuous-time discrete-space (CTDS) formulation for movement [HHA15]. The motivation for developing CTDS models comes from the way landscape covariates are typically recorded. Many commonly studied landscape features are available as raster images over a discrete space of pixels, with the value in each pixel representing a value aggregated over a non-zero area on the landscape. The resolution for these landscape covariates can vary considerably, and often it is the case that the granularity of the raster is noticeable compared to the scale of behavior. Thus, it makes sense to model an individual's decision-making process as a sequence of transitions from pixel to pixel. By modeling expected residence times as a function of the landscape characteristics of a given pixel, and transition probabilities between adjacent pixels as a function of both the

current and prospective pixel, CTDS approaches can characterize which landscape types animals prefer, and which landscape types facilitate/hinder rapid movement. Understanding these relationships can reveal landscape features that may act as potential corridors for long-distance movement, provide insight into the dynamics of disease spread, and help predict the impact of new human activity and development on a species' landscape use. Though not immediately obvious, it can be shown that CTDS models are examples of generalized linear models (GLMs) [HHA15]. GLMs are a well-established and theoretically rigorous class of statistical models readily implemented with commonly available statistical software.

In a CTDS model, a trajectory through a discrete grid is represented as a sequence of cell labels, g_i , and associated residence time, τ_i . Continuity of the trajectory requires that g_i and g_{i+1} be adjacent and $\tau_i > 0$. As with discrete-time HMMs, there is a misalignment that arises between observations of position, which are made in continuous space, and the discrete-space process model. Multiple imputation has been used to reconcile this mismatch analogous to the way it has been used in two-stage discrete-time approaches. Figure 5 shows portions of two of the realizations from Figure 2 (red, blue curves) through an artificially low-resolution bathymetry map. The red and blue continuous-space trajectories each have corresponding sequences of cells and residence times in the discrete space. Using multiple imputation allows for the propagation of uncertainty about the true sequence of visited cells and residence times.

Advanced Mechanistic Models for Animal Movement

HMMs for movement provide inference about hypothesized behavior states that may explain a large amount of variability in animal trajectories, but there are limitations to the behavior mechanism implied by a discrete set of latent behaviors. The meaning of the latent states is usually determined post hoc, and the interpretation becomes much less clear as the number of states grows beyond two or three. The incorporation of environmental covariates into a model for transition probabilities in (22) offers some explanation for why an animal is more likely to be in a particular state at any given time, but the effect of the covariate is limited to a local interpretation. If, for example, an animal changes from an undirected to directed type of behavior for the purpose of reaching a distant food source it knows about, it would not be possible to explain this action with local environmental characteristics.

In light of these limitations, HMMs can be considered a step in the mechanistic direction along the phenomenological continuum, but there are opportunities for further development. Recently, novel movement models

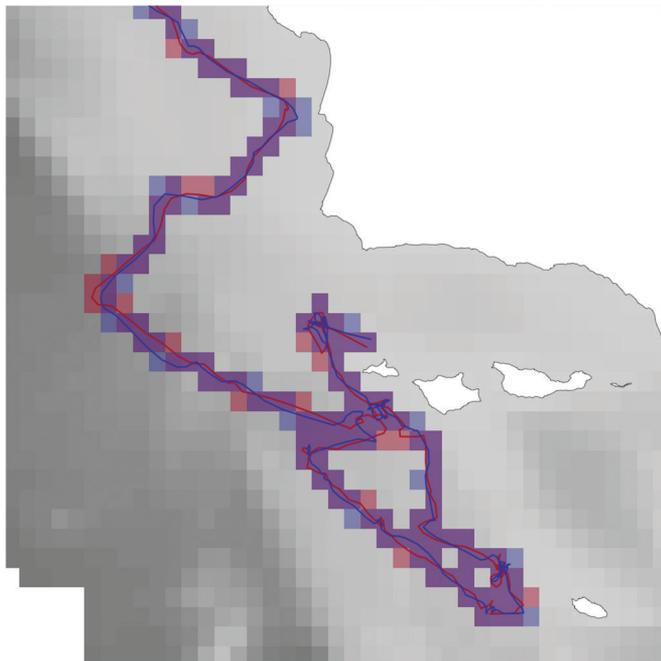


Figure 5. Cell sequences for two continuous-time continuous-space trajectories selected from the Johnson et al. [JLLD08] model shown in Figure 2. Purple cells represent cells visited by both trajectories.

have been proposed that attempt to apply clear a priori meaning to behavior states, and explain changes in behavior through physiology and memory of the landscape. The framework laid out in [HSM19] allows for behavior states based on attraction toward/repulsion from important landscape features. In addition, changes in behavior state are explained through a latent “charge” function that aggregates various physiological conditions such as hunger and thirst.

For example, the movement of a mountain lion (*Puma concolor*) in Colorado over the course of several weeks in the spring of 2017 was modeled in [HSM19] as a mixture of two behaviors: an undirected random walk, and a biased random walk with a direction determined by several landscape covariates including slope, elevation, ruggedness, and distance to a known prey kill area. By fitting the model to observations made approximately every three hours, it was concluded that the directed behavior was consistent with a need to return to the vicinity of the prey kill area, and the probability of entering the directed behavior state steadily increased as the animal spent time away from the prey kill area.

Models for new data. New types of data related to animal movement are now emerging thanks to continued technological advances. Accelerometers, which provide precise measurements of acceleration up to several times per second, are increasingly commonplace in animal movement studies. Coupled with relatively infrequent position

observations, accelerometers offer the potential for extremely precise position estimates that may obviate the need to consider measurement error.

Drones are another tool that can provide precise position information for many species, and without the need to affix transmitters to animals. While limited in their ability to reach remote locations, drones offer a relatively unintrusive way to study species such as dolphins that are difficult for humans to physically follow, but often spend time in the vicinity of human activity. A single drone is able to precisely track the positions of several interacting individuals in near-continuous time, which, coupled with an appropriate statistical model, could provide information about social structure and changes to social relationships in response to exogenous influences such as increased human interactions.

Methodological development for animal movement data is a vibrant and active field of ongoing research. New sources of data and new scientific questions demand novel statistical models and computationally efficient implementation strategies. In addition, new statistical methods can expose new information in legacy data, and synthesize historic and modern observations to provide insight into long-term trends in behavior. As human activity continues to expand interaction with previously isolated ecologies, and climate rapidly alters habitat around the globe, increasingly urgent scientific exploration is driving an immediate need for quantitative methods to study animal behavior.

References

- [BHIS16] Frances E. Buderman, Mevin B. Hooten, Jacob S. Ivan, and Tanya M. Shenk, *A functional model for characterizing long-distance movement behaviour*, *Methods in Ecology and Evolution* 7 (2016), no. 3, 264–273.
- [Cen] NOAA National Geophysical Data Center, 2003: *U.S. Coastal Relief Model Vol.6 - Southern California*, NOAA National Centers for Environmental Information. <https://doi.org/10.7289/V500001J>. Accessed 07/21/2020.
- [DdFG01] Arnaud Doucet, Nando de Freitas, and Neil Gordon (eds.), *Sequential Monte Carlo methods in practice*, *Statistics for Engineering and Information Science*, Springer-Verlag, New York, 2001. MR1847783
- [DK01] J. Durbin and S. J. Koopman, *Time series analysis by state space methods*, *Oxford Statistical Science Series*, vol. 24, Oxford University Press, Oxford, 2001. MR1856951
- [GCS⁺14] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin, *Bayesian data analysis*, 3rd ed., *Texts in Statistical Science Series*, CRC Press, Boca Raton, FL, 2014. MR3235677
- [HHA15] Ephraim M. Hanks, Mevin B. Hooten, and Mat W. Allredge, *Continuous-time discrete-space models for animal movement*, *Ann. Appl. Stat.* 9 (2015), no. 1, 145–165, DOI 10.1214/14-AOAS803. MR3341111

- [HT86] Trevor Hastie and Robert Tibshirani, *Generalized additive models*, *Statist. Sci.* **1** (1986), no. 3, 297–318. With discussion. MR858512
- [HJMM17] Mevin B. Hooten, Devin S. Johnson, Brett T. McClintock, and Juan Manuel Morales, *Animal movement: Statistical models for telemetry data*, Chapman and Hall/CRC, Boca Raton, Florida, USA, 2017.
- [HSM19] Mevin B. Hooten, Henry R. Scharf, and Juan Manuel Morales, *Running on empty: Recharge dynamics from animal movement data*, *Ecology Letters* **22** (2019), 377–389.
- [IWF+20] Ladd M. Irvine, Martha H. Winsor, Tomas M. Follett, Bruce R. Mate, and Daniel M. Palacios, *Blue and fin whales Southern California 2014–2015-Argos data*, Movebank Data Repos (2020).
- [JLLD08] Devin S. Johnson, Joshua M. London, Mary-Anne Lea, and John W. Durban, *Continuous-time correlated random walk model for animal telemetry data*, *Ecology* **89** (2008), no. 5, 1208–1215.
- [Kal60] R. E. Kalman, *A new approach to linear filtering and prediction problems*, *Trans. ASME Ser. D. J. Basic Engrg.* **82** (1960), no. 1, 35–45. MR3931993
- [KCJW15] Roland Kays, Margaret C. Crofoot, Walter Jetz, and Martin Wikelski, *Terrestrial animal tracking as an eye on life and planet*, *Science* **348** (2015), no. 6240.
- [LHB+14] Roland Langrock, J. Grant C. Hopcraft, Paul G. Blackwell, Victoria Goodall, Ruth King, Mu Niu, Toby A. Patterson, Martin W. Pedersen, Anna Skarin, and Robert S. Schick, *Modelling group dynamic animal movement*, *Methods in Ecology and Evolution* **5** (2014), no. 2, 190–199, available at 1308.5850.
- [MB19] Théo Michelot and Paul G. Blackwell, *State-switching continuous-time correlated random walks*, *Methods in Ecology and Evolution* **10** (2019), no. 5, 637–649.
- [Nat08] Ran Nathan, *An emerging movement ecology paradigm*, *Proc. Natl. Acad. Sci. USA* **105** (2008), no. 49, 19050–19051.
- [PTW+08] Toby A. Patterson, Len Thomas, Chris Wilcox, Otso Ovaskainen, and Jason Matthiopoulos, *State-space models of individual animal movement*, *Trends in Ecology and Evolution* **23** (2008), no. 2, 87–94.
- [Rub96] Donald B. Rubin, *Multiple imputation after 18+ years*, *Journal of the American Statistical Association* **91** (1996), no. 434, 473–489.
- [SHF+16] Henry R. Scharf, Mevin B. Hooten, Bailey K. Fosdick, Devin S. Johnson, Josh M. London, and John W. Durban, *Dynamic social networks based on movement*, *Ann. Appl. Stat.* **10** (2016), no. 4, 2182–2202, DOI 10.1214/16-AOAS970. MR3592053
- [SHJ17] Henry Scharf, Mevin B. Hooten, and Devin S. Johnson, *Imputation approaches for animal movement modeling*, *J. Agric. Biol. Environ. Stat.* **22** (2017), no. 3, 335–352, DOI 10.1007/s13253-017-0294-5. MR3692468
- [Woo20] Simon N. Wood, *Inference and computation with generalized additive models and their extensions*, *TEST* **29** (2020), no. 2, 307–339, DOI 10.1007/s11749-020-00711-5. MR4095042

- [ZML16] Walter Zucchini, Iain L. MacDonald, and Roland Langrock, *Hidden Markov models for time series: An introduction using R*, 2nd ed., *Monographs on Statistics and Applied Probability*, vol. 150, CRC Press, Boca Raton, FL, 2016. MR3618333



Henry Scharf

Credits

Opening image, Figures 1–5, and author photo are courtesy of the author.

Mathematical and
Statistical Sciences
Annual Survey

DATA ON THE COMMUNITY

DOCTORAL RECIPIENTS
New PhD graduates, their employment plans,
demographics, and starting salaries

DOCTORAL DEGREES & THESIS TITLES
PhD graduates, their thesis titles, and where they earned their degrees

FACULTY SALARIES
By rank and employment status

RECRUITMENT & HIRING
The academic job market

DEPARTMENTAL PROFILE
The number of—faculty, their employment statuses
and demographics; course enrollments; graduate students;
masters and bachelors degrees awarded

[www.ams.org/
annual-survey](http://www.ams.org/annual-survey)

Sponsored by:
AMS | ASA | IMS | MAA | SIAM