STAT 548 Report 4

# "When to be discrete: the importance of time formulation in understanding animal movement"

McClintock, Johnson, Hooten, Ver Hoef, Morales
October 2014

Submitted to
Dr. James V. Zidek
March 16, 2018

Saifuddin Syed
85258144

# Contents

# 1 Introduction

*"All models are wrong, but some are useful."*

*-George Box, 1976*

We are surrounded by a web of complex systems that govern why the world around us works the way it does. Applied scientists/mathematicians are given with the arduous task of observing these seemingly chaotic phenomenon with the hopes of finding some structure and make inferences. This is done systematically by collecting data and using it to develop models that aim to answer questions we have about our system under the umbrella of the scientific method. As technology advances, the quality and quantity of data we have access to improves, which in turn, necessitates the models are refined accordingly to better approximate the phenomenon of interest.

As much as we would like to think otherwise, designing a good model is an art. We have a great deal of flexibility with regards to the properties we want our models to have or not have. As statistician George Box, famously stated "all models are wrong, but some are useful", we always need to make sacrifices between accuracy in favour of tractability and interpretability, but this does not mean the design choices we make can still be very useful to gain intuition about our system. It is the task of a statistician to strike that balance between developing a model that accurately allows us to make inferences about our system while working within the limitations of our data, mathematical tools, and computational power. The design choices we make are fundamental, and it is important to be critical and understand their strengths and limitations.

One such task is comes from ecology, and the study of tracking mammal movement. Mammal tracking is a complicated phenomenon with a variety of spatial and temporal factors that govern an animal moves. In this report, we will focus our attention to comparing and contrasting the design choice of using a discrete or continuous time model for mammal tracking. Specifically we will analyse the case study on Seal tracking tracking data.

We will first give an overview of the problem, discuss how it can be formulated as a discrete and continuous time problem. We will then compare and contrast these models, and discuss possible extensions.

# 2 Characterization of the Movement Process

Most analysis of animal location data is broken into two distinct components: the movement process and the observation model process. The movement process $X = (X)_t$ is the assumed underlying unobserved state of the system of interest at time $t$ parametrized by hyper parameters $\theta$. The observational process $Y_t$ interprets the data $\mathcal{D}$ as observed noisy measurements of $X_t$. The class of models we will be working with are commonly referred to as hierarchical models or state-space models (SSM). When the movement process is Markovian, we refer this as a hidden Markov model (HMM).

The effectiveness of state space model come from the movement process, which we have a lot of freedom in constructing. We will demonstrate this by focussing our attention to the tracking of North Fur seals, noting that the ideas can be adapted to modelling movement of other mammals.

## 2.1 Seal tracking data

For the purpose of this case study we will be developing continuous and discrete time space models to a northern fur seal track in Pribilof Islands of Alaska as during a 10 day foraging trip. The data was attained by equipping a nursing female with a satellite tag, which has Fastloc GPS, and time-depth recoding capabilities. Fastloc GPS are designed to specifically to track species that only surface briefly by quickly take snapshots of the radio signals produced by overhead GPS satellites [fas17]. This is ideal for seal tracking as they are often foraging underwater and may only resurface for short periods of time.

The diving depth data is collected continuously and is summarized every hour over a 10 day period (242 data points), and the raw location data consists of 241 data points with over 40% of the time steps containing no observed location due to foraging dives.

We are making an assumption that the seal are always in one of three latent (unobserved) states: foraging, transit, resting, which we define as follows:

- *Foraging (F):* movement that is characterized as area restricted searches and foraging dives. A foraging dive is a dive that has a max depth of more than 5m and atleast 5 changes in vertical direction.
- *Transit (T):* movement that is predominantly travelling with little to know foraging dives.
- *Resting (R):* Movements that are not foraging or transit. We will assume that no foraging occurs while resting.

Using the location and diving activity data, our goal is to characterize the latent behaviour of the seals during their foraging trip. We should expect to see low speeds during resting, low to moderate speed when foraging, and high speeds for transit states. We also expect to see high amounts of directional persistence during transit and little to none when resting and foraging.

In order for our movement model to incorporate the different types of movement, we want the movement process to have postitive correlations in the direction of travel between time steps. This will also ensure directional persistency, which we want for the transit and foraging states. We would also like to define out model in such a way that it can have a Bayesian interpretation.

# 3 Discrete Time Formulation

We will first begin by formulating how to formulate a measurement model from our seal data, and then we will develop a discrete time model following framework laid out in [MKT$^+$12].

## 3.1 Measurement Model

When developing a discrete model, we need to work with fixed time steps, but since the location data $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_{t,i}, y_{t,i}) : t = 1, \ldots, N, i = 1, \ldots, k_t\}$ is temporally irregular we need to be a bit careful with how we handle it. First off, we will work a temporal units of 1 hour, so for $t = 1, \dots, N$, we denote the position of the seals at time $t$ hours as $\mu_t^{dis} = (X_t, Y_t)$. We can extend $\mu_s^{dis}$ for $t < s < t+1$, by linear interpolation. Within time $[t-1, t]$, we have $i = 1, \dots, k_t$ observations $(x_{t,i}, y_{t,i})$ which take place at time $j_{t,i} \in [0, 1)$ hours after time $t-1$. We will assume that $(x_{t,i}, y_{t,i})$ noise measurements of $\mu_{t-1+j_{t,i}}$, or equivalently,

$$x_{t,i} = (1 - j_{t,i})X_{t-1} + j_{t,i}X_t + \varepsilon_{x_{t,i}} \tag{1}$$

$$y_{t,i} = (1 - j_{t,i})Y_{t-1} + j_{t,i}Y_t + \varepsilon_{y_{t,i}}, \tag{2}$$

where $[\varepsilon_{x_{t,i}}] = \mathcal{N}(0, \sigma_x^2)$, and $[\varepsilon_{y_{t,i}}] = \mathcal{N}(0, \sigma_y^2)$are independent and represent the measurement error from the GPS tracking [1]. This implies we are assuming that the measurement error in the $x$-coordinate is independent of the measurement in the $y$-coordinate and furthermore the measurement error is independent of time.

At time $t$, denote the behaviour state of our process as $Z_t$, where $Z_t \in \{F, T, R\}$ as defined in Section 2.1. We also want to incorporate dive data $\boldsymbol{\delta} = \{\delta_t : t = 1, \dots, N\}$ which is the number of foraging dives in the hour time span into our measurement model. Given $Z_t = z$, for $z \in \{F, T, R\}$, we will assume that

$$[\delta_t | \boldsymbol{\lambda}, Z_t = z] = Poisson(\lambda_z), \tag{3}$$

where $\lambda_F > \lambda_T$, and $\lambda_R = 0$ to indicate more that one expects more forage dive attempts when foraging as opposed to the opportunistic foraging that occurs in transit. $\lambda_R = 0$ forces the condition that no diving occurs when resting. We will futher assume that conditioned on $Z_t$, $\delta_t$ is independent of the position.

This is an okay assumption as the Poisson distribution assumes a certain independence between dive attempts. Perhaps it may be better to use a geometric distribution to model the first time a dive was unsuccessful. It is reasonable to assume that if a dive was successful, a seal would be move likely to reattempt that dive, and one would stop diving if no food was caught.

Thus our discrete measurement model can be parametrized by $(\sigma_x^2, \sigma_y^2, \boldsymbol{\lambda})$.

## 3.2 Movement Model

We will now create a model for $\mu_t^{dis} = (X_t, Y_t)$ for $t = 1, \dots, N$. Given an initial $\mu_0^{dis} = (X_0, Y_0)$ we wish to iteratively define how to move from position $\mu_t^{dis}$ to $\mu_{t+1}^{dis}$. Let $V_t^{dis} = \mu_t^{dis} - \mu_{t-1}^{dis} = (V_{x,t}^{dis}, V_{y,t}^{dis})$ be the discrete velocity process so that we have

$$\mu_t^{dis} = \sum_{s=0}^{t} V_s^{dis}.$$

It will be more natural to describing $V_t^{dis}$ in polar coordinates $(\phi_t, S_t)$ as opposed to Cartesian coordinates. Specifically, define the step size $S_t = \|V_t^{dis}\|$ and the bearing angle $\phi_t$. Ie

$$S_t = \sqrt{(V_{x,t}^{dis})^2 + (V_{y,t}^{dis})^2} = \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2},$$

[1] We are following the notation in the paper [MJH+14], $[X]$ represents the density of the random variable $X$ inside the brackets.

and,

$$\phi_t = \arctan\left(\frac{V_{y,t}^{dis}}{V_{x,t}^{dis}}\right) = \arctan\left(\frac{Y_t - Y_{t-1}}{X_t - X_{t-1}}\right)$$

Conditional on the behaviour state $Z_t$, we will assume that $S_t$ and $\phi_t$ are independent. We now have flexibility in how we chose to specify a family of distributions for $X_t^{dis} = (S_t, \phi_t, Z_t)$, which we will discuss now. We will refer to our discrete movement model $\mathbf{X}^{dis} = (\mathbf{S}, \boldsymbol{\phi}, \mathbf{Z})$.

### 3.2.1 Step Size

Following the work of [MKT+12], we will assume,

$$[S_t|\mathbf{a}, \mathbf{b}, Z_t = z] = Weibull(a_z, b_z) = \frac{b_z}{a_z}\left(\frac{S_t}{a_z}\right)^{b_z - 1} \exp\left[-\left(\frac{S_t}{a_z}\right)^{b_z}\right], \tag{4}$$

where $z \in \{F, T, R\}$ and $a_z, b_z > 0$ are scale and shape parameters. Note that given $S_t$ is fully determined by the latent behaviour as is independent of previous step sizes. Weibull distribution has the ability to assume the characteristics of many different types of distribution and varies dramatically with $b_z$. When $b_z < 1$, the Weibull distribution has heavy tails, when $b_z = 1$, we get an exponential distribution, for $1 < b_z$, the Weibull distribution resembles a bell curve at with skewness governed by $b_z$. It even approximate a normal distribution for $b_z = 3.4$. This versatility makes it robust to the different possible behaviours we may want from $S_t$.

A possibly more suitable alternative for this specific paper would have been the Rice distribution $R(\mu, \sigma^2)$, which is the distribution of the distance from the origin of a bivariate Gaussian random vector with covariance $\Sigma = \sigma^2 I$. This is because it arrives naturally as the distribution of the step size in the continuous time model, and would have allowed for a better comparison of the two methods.

### 3.2.2 Bearing angle

We need to chose our distribution for $\phi_t$ carefully as $\phi_t$ governs how correlated the direction of the movement is between time steps. If $\phi_t$ was uniform on $[0, 2\pi]$, then it would mean that conditional on the behaviour state, the movement is process is a random walk and has no direction correlation/persistence. In general we want our directional movement to be biased towards $\phi_{t-1}$ as that imply directional persistence. There are many distributions on the circle that concentrate around $\phi_{t-1}$, in particular every distribution $f(x)$ on $\mathbb{R}$ induces a "wrapped" distribution $f_w(x)$ on the circle $[0, 2\pi]$ by identifying $x$ with $x + 2k\pi$, $k \in \mathbb{Z}$.

$$f_w(x) = \sum_{k=-\infty}^{\infty} f(x + 2k\pi)$$

One natural example would be a wrapped Gaussian, but due to the infinite sum, it is not an nice to work with. We model the bearing angle $\theta_t$ using the wrapped Cauchy distribution centered at $\theta_{t-1}$, as the sum can be simplified to a nice formula using some complex analysis to the following.

$$[\phi_t|\phi_{t-1}, \boldsymbol{\rho}, Z_t = z] = wCauchy(\phi_{t-1}, \rho_z) = \frac{1-\rho_z^2}{2\pi[1+\rho_z^2 - 2\rho_z\cos(\phi_t - \phi_{t-1})]}, \tag{5}$$

Where $\rho_z \in [0, 1)$ is a hyper-parameter governing the dispersion of the bearing angle. As $\rho_z \to 0$, we have the wrapped Cauchy begins to approach a uniform distribution on the circle since $[\phi_t|\phi_{t-1}, \boldsymbol{\rho}, Z_t = z] \to \frac{1}{2\pi}$. To see what happens when $\rho_z \to 1^-$, note $[\phi_t|\phi_{t-1}, \boldsymbol{\rho}, Z_t = z] \to 0$ as $|\rho_z| \to 1$ for $\phi_t \neq \phi_{t-1}$. If $\phi_t = \phi_{t-1}$, then

$$[\phi_{t-1}|\phi_{t-1}, \boldsymbol{\rho}, Z_t = z] = \frac{1-\rho_z^2}{2\pi(1-\rho_z)^2)} = \frac{1}{2\pi}\frac{1+\rho_z}{1-\rho_z} \xrightarrow{\rho_z \to 1^-} \infty$$

and thus $[\phi_t|\phi_{t-1}, \boldsymbol{\rho}, Z_t = z] \to \delta_{\phi_{t-1}}(\phi_t)$.

We are leaving out the formal details as they are not very insightful, mainly because there is nothing particularly special about using the wrapped Cauchy than the mathematically convenience that comes with a closed formula. This was the reasoning used in [MKT+12], where this design choice was proposed. In the continuous time model as we will soon see, the bearing has a Von Mises distribution, which is commonly referred to as the circle analogue of the Normal distribution [MEP01]. Thus using a Von Mises distribution for the bearing, it would have allowed for a more direct comparison of the results between continuous and discrete time models.

### 3.2.3 Latent Behaviour Model

Finally we need to model how our latent behaviour state $Z_t$ changes between time steps. We will assume that at each time step, $Z_t$ is a stationary and in fact a Markov process with transition kernel $\boldsymbol{\psi}$, i.e. $P(Z_t = z|Z_{t-1} = z') = \psi_{z',z}$, or to continue with our notation above,

$$[Z_t|\boldsymbol{\psi}, Z_{t-1} = z] = Categorical(\psi_{z,F}, \psi_{z,T}, \psi_{z,R}). \tag{6}$$

This is definitely a reasonable assumption to make, as the the current state you are in is far more relavent to the furture decision to forage, more, or rest for a seal. It is reasonable to assume that incorporating $Z_{t-2}, \ldots, Z_{t-k}$ via an auto regressive time series model, might lead to slightly more accurate predictions. However in the context of seal tracking, one would assume the auto-correlation function $f(k) = \mathrm{Cov}(Z_t, Z_{t-k})$ to decrease rapidly as $k$ increases, and by incorporating too many previous time steps, could possibly lead to over-fitting.

Overall the Markovian assumption is a reasonable one and the gains we get in terms of model accuracy are overshadowed by the mathematical difficulties that will arise by losing the Markov property.

## 3.3 Posterior Model

Finally, now that we have specified our measurement and movement model, we can now compute the posterior of our hyper-parameters

$$\Theta = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \sigma_x^2, \sigma_y^2, (X_0, Y_0)\}.$$

We will now compute the joint posterior density of $\Theta$ and our movement process $\mathbf{X}^{dis} = (\mathbf{S}, \boldsymbol{\phi}, \mathbf{Z})$, and data $\mathcal{D} = \{\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}\}$. Let $[\Theta]$ denote the prior of our hyper-parameters, which we will discuss below. First let us compute the joint posterior. We first use Bayes theorem, and the product rule.

$$\begin{aligned}
[\Theta, \mathbf{X}^{dis}|\mathcal{D}] &\propto [\mathbf{X}^{dis}, \mathcal{D}|\Theta][\Theta] \\
&= [\mathbf{S}, \boldsymbol{\phi}, \mathbf{Z}, \mathbf{x}, \mathbf{y}, \boldsymbol{\delta}|\Theta][\Theta] \\
&= [\mathbf{S}, \boldsymbol{\phi}, \mathbf{x}, \mathbf{y}, \boldsymbol{\delta}|\Theta, \mathbf{Z}][\mathbf{Z}|\Theta][\Theta] \\
&= [\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}|\Theta, \mathbf{Z}, \mathbf{S}, \boldsymbol{\phi}][\mathbf{S}, \boldsymbol{\phi}|\Theta, \mathbf{Z}][\mathbf{Z}|\Theta][\Theta] \\
&= [\boldsymbol{\delta}|\Theta, \mathbf{Z}, \mathbf{S}, \boldsymbol{\phi}][\mathbf{x}, \mathbf{y}|\Theta, \mathbf{Z}, \mathbf{S}, \boldsymbol{\phi}] \\
&\quad \times [\mathbf{S}|\Theta, \mathbf{Z}][\boldsymbol{\phi}|\Theta, \mathbf{Z}][\mathbf{Z}|\Theta][\Theta].
\end{aligned}$$

The last line used the assumption that conditioned on $Z$, that $\boldsymbol{\delta}$ is independent of the position data $\mathbf{x}, \mathbf{y}$, and $\mathbf{S}$ is independent of $\boldsymbol{\phi}$. Combining (1),(3),(4),(5),(6), we get,

$$\begin{aligned}
[\Theta, \mathbf{X}^{dis}|\mathcal{D}] \propto [\Theta] \prod_{t=1}^{N} \{ [\delta_t|\boldsymbol{\lambda}, Z_t] \prod_{i=1}^{k_t} [x_{t,i}, y_{t,i}|\sigma_x^2, \sigma_y^2, X_0, Y_0, \phi_{1:t}, S_{1:t}] \\
\times [S_t|\mathbf{a}, \mathbf{b}, Z_t][\phi_t|\phi_{t-1}, \boldsymbol{\rho}, Z_t][Z_t|\boldsymbol{\psi}, Z_{t-1}] \}
\end{aligned}$$

For our prior, we have

$$[\Theta] = [\mathbf{a}][\mathbf{b}][\boldsymbol{\rho}][\boldsymbol{\lambda}][\boldsymbol{\psi}][\boldsymbol{\lambda}][\sigma_x^2][\sigma_y^2][X_0, Y_0]$$

Where we have,

$$\begin{aligned}
[a_z] &= Uniform(0, 10800) & z &\in \{F, T, R\} \\
[b_z] &= Uniform(0, 3) & z &\in \{F, T, R\} \\
[\rho_z] &= Uniform(0, 1) & z &\in \{F, T, R\} \\
[\lambda_z] &= \Gamma(0.1, 0.1) & z &\in \{F, T\} \\
[\boldsymbol{\psi}_z] &= Dirichelet(1, 1, 1) & z &\in \{F, T, R\} \\
[\sigma_x^2] &= \Gamma^{-1}(0.1, 0.1) \\
[\sigma_y^2] &= \Gamma^{-1}(0.1, 0.1),
\end{aligned}$$

and $[X_0, Y_0]$ is defined uniformly over the Bering Sea. These weakly informative priors are chosen as they assume very little information about the system and are computationally easy to work with, since they are either flat or conjugate to their respective families.

# 4 Continuous Time Formulation

We will now proceed to develop a continuous time model analogous to the discrete time model above. We will be modelling the continuous time latent position process $\mu_t^{cts}$, by the continuous time correlated random walk (CTCRW) model of Johnson in [JLLD08]. Analogous to the discrete setting, CTCRW model uses a on Ornstein-Uhlenbeck (OU) process to stochastically describe the velocity $V_t^{cts}$, and using that to derive the position process $\mu_t^{cts}$ by integrating $V_t^{cts}$. Let us first describe the measurement model, and then we will construct the movement model.

## 4.1 Measurement, Latent Space Model

Our measurement model in the continuous setting will be very similar to that of the discrete setting with some mild simplifications. To allow for comparisons to the discrete time model latent space model, we will still assume a temporal resolution of 1 hour. We will assume we have the same location data $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_{t,i}, y_{t,i}) : t = 1, \ldots, N, i = 1, \ldots, k_t\}$ as in the discrete setting where the observation $(x_{t,i}, y_{t,i})$ occurs at time $t < t_i < t + 1$ and $t_i < t_{i+1}$. Unlike the discrete-time setting where the measurement model for the location data required linear interpolation of the latent position process $\mu^{dis}$, in the continuous case out we can avoid that messiness. Let $\mu_{t_i}^{cts} = (X_{t_i}, Y_{t_i})$, then

$$x_{t,i} = X_{t_i} + \varepsilon_{x,t_i} \tag{7}$$

$$y_{t,i} = Y_{t_i} + \varepsilon_{y,t_i}, \tag{8}$$

Where $\varepsilon_{c,t_i} \sim \mathcal{N}(0, \tau^2)$ are iid. This is slightly different than the discrete setting where we assumed that each coordinate had a different movement measurement error. This was the design choice made in [MJH+14], which I personally disagree with. To allow for a better comparison to the discrete-time model, we should have assumed different measurement error in each coordinate, as computationally it is not expensive.

The measurement model for the dive date $\boldsymbol{\delta}$ is identical to the discrete-setting, and we will use $[\delta_t|\boldsymbol{\lambda}, Z_t = z]$ as in (3).

Finally to allow for direct comparison to the discrete-time model, we will assume that the latent state model $Z_t$ is unchanged from the discrete setting, including the fact that it only changes for $t = 1, \ldots, N$. We will also assume that $[Z_t|\boldsymbol{\psi}, Z_{t-1}]$ is the same as (6).

Personally, I think was a poor choice made by the authors, as this imposed the discrete-time restrictions of the continuous-time model, where hours $t = 1, 2, \ldots, N$ are not the natural time of intererst, but rather we care about $t_i$. The continuous time analogue would have been to model the transition between state $z \to z'$ via an exponential distribution with rate $\psi_{z,z'}$. This would have retained the same interpretation of $\boldsymbol{\psi}$ from the discrete-time model, as well have eliminated the some of the hassle that comes with dealing having to deal with observations that occur at time $t_i < t < t_{i+1}$ where $t$ is an integer.

We are almost ready to introduce the CTCRW, but first it will be worth formally introducing the OU process.

## 4.2 Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck process is a continuous time stationary, Markov, Gaussian process, characterized by the stochastic differential equation,

$$dV_t = -\beta(V_t - \gamma)dt + \sigma dW_t, \tag{9}$$

where $\beta, \sigma > 0$ are the autocorellation parameter and standard deviation respectively and $\gamma \in \mathbb{R}$ is the mean drift and $W_t$ is a standard Brownian motion. Note that if $V_t > \gamma$ (similarly $V_t < \gamma$), then the drift component is negative (positive) and will have a tendency to decrease (increase) towards $\gamma$. Thus the solution to (9) has a tendency to revert to the $\gamma$. We can solve (9) explicitly by defining $Y_t = \exp^{\beta t}(V_t - \gamma)$. Ito's lemma gives us the following SDE for $Y$,

$$\begin{aligned} dY_t &= \beta e^{\beta t}(V_t - \gamma)dt + e^{\beta t}dV_t \\ &= \beta e^{\beta t}(V_t - \gamma)dt + e^{\beta t}(-\beta(V_t - \gamma)dt + \sigma dW_t) \\ &= \sigma e^{\beta t}dW_t \end{aligned}$$

By integrating both sides from $s$ to $t$ we get,

$$Y_t = Y_s + \sigma \int_s^t e^{\beta u}dW_u.$$

Substituting $Y$ and simplifying leads to our solution to (9).

$$\begin{aligned} e^{\beta t}(V_t - \gamma) &= e^{\beta s}(V_s - \gamma) + \sigma \int_s^t e^{\beta u}dW_u. \\ V_t &= \gamma + e^{-\beta(t-s)}(V_s - \gamma) + \sigma e^{-\beta t}\int_s^t e^{\beta u}dW_u. \end{aligned} \tag{10}$$

To compute the distribution of the stochastic integral the following result from stochastic analysis will be helpful.

**Lemma 4.1.** *For $f(t), g(t)$ be deterministic functions, and $W_t$ be a standard Weiner Process, then,*

1. $\displaystyle\int_s^t f(u)dW_u \sim \mathcal{N}\left(0, \int_s^t f(u)^2 du\right)$.
2. $\displaystyle\mathrm{Cov}\left(\int_s^t f(u)dW_u, \int_s^t g(u)dW_u\right) = \int_s^t f(u)g(u)du$

   Using this we get that

$$\sigma e^{-\beta t}\int_s^t e^{\beta u}dW_u \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right).$$

To see how this related to the auto-regressive model $AR(1)$, let us fix $\Delta t$, and define $Z_k = V_{k\Delta}$. Then (10) implies that,

$$\begin{aligned} Z_{k+1} &= \gamma + e^{-\beta\Delta t}(Z_k - \gamma) + \varepsilon_k \\ &= a + bZ_k + \varepsilon_k \end{aligned}$$

where $b = e^{-\beta\Delta t}$, $a = (1-b)\gamma$ and

$$\varepsilon_k \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(t-s)})\right).$$

Thus we can see that the sampled OU process is the solution to an $AR(1)$ model which implies the OU process is the continuous time analogue of the an $AR(1)$ model.

### 4.2.1 Integrated OU Process

Suppose that when $\gamma = 0$, in which case (10) reduces to,

$$V_{t+\Delta} = e^{-\beta\Delta}V_t + \sigma \int_t^{t+\Delta} e^{\beta(t+\Delta-u)} dW_s$$
$$\equiv e^{-\beta\Delta}V_t + \varepsilon_V(\Delta), \tag{11}$$

where

$$\varepsilon_V(\Delta) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\Delta})\right).$$

Let us look at the integrated OU process $X_t = X_0 + \int_0^t V_s ds$ be defined by the differential equation,

$$dX_t = V_t dt. \tag{12}$$

Note that we can substitute (12) into (9) to get,

$$dV_t = -\beta dX_t + \sigma dW_t,$$

Integrating both sides from $t$ to $t + \Delta$,

$$V_{t+\Delta} - V_t = -\beta(X_{t+\Delta} - X_t) + \sigma \int_t^{t+\Delta} dW_s.$$

Rearranging terms, and using (11), we get,

$$X_{t+\Delta} = X_t - \frac{1}{\beta}(V_{t+\Delta} - V_t) + \frac{\sigma}{\beta} \int_s^t dW_s$$
$$= X_t + \frac{1 - e^{-\beta\Delta}}{\beta} V_t + \frac{\sigma}{\beta} \int_t^{t+\Delta} (e^{-\beta(t+\Delta-s)} + 1) dW_s$$
$$\equiv X_t + \frac{1 - e^{-\beta\Delta}}{\beta} V_t + \varepsilon_X(\Delta) \tag{13}$$

Where using Lemma 4.1, we have

$$\varepsilon_X(\Delta) \sim \mathcal{N}\left(0, \frac{\sigma^2}{\beta^2}\left(\Delta - \frac{2}{\beta}(1 - e^{-\beta\Delta}) + \frac{1}{2\beta}(1 - e^{-2\beta\Delta})\right)\right).$$

If $\boldsymbol{\alpha}_t = (X_t, V_t)$, then (11), and (13) implies, that

$$\boldsymbol{\alpha}_{t+\Delta} = \mathbf{T}_\Delta \boldsymbol{\alpha}_t + \boldsymbol{\eta}_\Delta,$$

where

$$\mathbf{T}_\Delta = \begin{bmatrix} 1 & \frac{1-e^{-\beta\Delta}}{\beta} \\ 0 & e^{-\beta\Delta} \end{bmatrix}, \tag{14}$$

$$\boldsymbol{\eta}_\Delta \sim \mathrm{N}(0, \mathbf{Q}_\Delta), \tag{15}$$

where, use Lemma 4.1 to get,

$$\begin{aligned}\mathbf{Q}_\Delta &= \begin{bmatrix} \mathrm{Var}(\varepsilon_X(\Delta)) & \mathrm{Cov}(\varepsilon_X(\Delta), \varepsilon_V(\Delta)) \\ \mathrm{Cov}(\varepsilon_X(\Delta), \varepsilon_V(\Delta)) & \mathrm{Var}(\varepsilon_V(\Delta)) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma^2}{\beta^2}\left(\Delta - \frac{2}{\beta}(1-e^{-\beta\Delta}) + \frac{1}{2\beta}(1-e^{-2\beta\Delta})\right) & \frac{\sigma^2}{2\beta^2}(1-2e^{\beta\Delta}+e^{-2\beta\Delta}) \\ \frac{\sigma^2}{2\beta^2}(1-2e^{\beta\Delta}+e^{-2\beta\Delta}) & \frac{\sigma^2}{2\beta}(1-e^{-2\beta\Delta}) \end{bmatrix}.\end{aligned}$$

## 4.3 Continuous-time Correlated Random Walk

Similar to the discrete-time model, we want to define the position process in such a way that we maintain correlation between movements, and so that we have directional persistency. We achieved this in the discrete-time setting by forcing the bearing angle in the $\phi_t$ in the discrete velocity process $V_t^{dis}$ to be be centered around $\phi_{t-1}$. Given the latent state $Z_t = z$, we have that the position process $\mu_t = \sum_{s=1}^t V_s$ is performing a correlated random walk with a tendency to go towards the previous direction of movement. This is precisely what the OU process aims to do in continuous time, and thus we will use it to define our velocity process $V_t^{cts}$. We will then define our continuous time position process $\mu_t^{cts} = \mu_0 + \int_0^t V_s^{cts} ds$, where $\mu_0 = (X_0, Y_0)$.

Given our latent state $Z_t = z$, we define $V_t^{cts} = (V_{x,t}^{cts}, V_{y,t}^{cts})$, be the velocity process for our movement mode, where $V_{x,t}^{cts}, V_{y,t}^{cts}$ are both independent OU processes with autocorrelation $\beta_z$ and variance $\sigma_z$ and zero drift ($\gamma = 0$). It was argued in [JLLD08], that the two components should be independent, as a correlation would result in strange movement patterns and unrealistic switching of directions.

We can now define our position process $\mu_t^{cts} = (X_t, Y_t)$ by the differential equations,

$$\begin{aligned} dX_t &= V_{x,t}^{cts} dt \\ dY_t &= V_{y,t}^{cts} dt. \end{aligned}$$

Let $\boldsymbol{\alpha}_t = (\alpha_{x,t}, \alpha_{y,t})$, where $\alpha_{x,t} = (\mu_{x,t}^{cts}, V_{x,t}^{cts})$, and $\alpha_{y,t} = (\mu_{y,t}^{cts}, V_{y,t}^{cts})$. Using the analysis of Section 4.2.1, we have for $c \in \{x, y\}$

$$\alpha_{c,t_{i+1}} = T_{z,\Delta_i} \alpha_{c,t_i} + \eta_{z,\Delta_i},$$

where $\Delta_i = t_{i+1} - t_i$, and $T_{z,\Delta}$, $\eta_{z,\Delta_i}$ were defined by (14), and (15).

Finally will add the constraint to our model that $\beta_R > \beta_F > \beta_T$ and $\sigma_R < \sigma_F < \sigma_T$. Since we expect less auto correlation and volatility in the velocity when moving versus resting.

Thus movement model $(\alpha_t)$ is parametrized by $\boldsymbol{\psi}, \boldsymbol{\beta}, \boldsymbol{\sigma}$, and the measurement model is parametrized by $(\tau^2, \boldsymbol{\lambda})$. Define out parameters $\Theta = \{\boldsymbol{\psi}, \boldsymbol{\beta}, \tau^2, \boldsymbol{\lambda}\}$, and our data as $\mathcal{D} = \{\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}\}$.

Since $\boldsymbol{\alpha}$ is a linear Gaussian model, we can sample form the posterior using a Kalman Filter without needing to sample $\boldsymbol{\alpha}$. We can now compute the posterior for the continuous time model:

$$
\begin{aligned}
[\Theta, \mathbf{Z}|\mathcal{D}] &\propto [\mathcal{D}, \mathbf{Z}|\Theta][\Theta] \\
&= [\mathcal{D}|\Theta, \mathbf{Z}][\mathbf{Z}|\Theta][\Theta] \\
&= [\boldsymbol{\delta}|\Theta, \mathbf{Z}][\boldsymbol{x}, \mathbf{y}|\Theta, \mathbf{Z}][\mathbf{Z}|\Theta][\Theta] \\
&= [\boldsymbol{\delta}|\boldsymbol{\lambda}, \mathbf{Z}][\boldsymbol{x}, \mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\sigma}, \tau^2, \mathbf{Z}][\mathbf{Z}|\boldsymbol{\psi}][\Theta] \\
&= [\Theta] \prod_{i=1}^{N} \{[\delta_t|\boldsymbol{\lambda}, Z_t] \prod_{i=1}^{k_t} \{[x_{t,i}, y_{t,i}|\beta_t, \boldsymbol{\sigma}, \tau^2, Z_t]\}[Z_t|\boldsymbol{\psi}, Z_{t-1}]\} \\
&= [\boldsymbol{\beta}][\boldsymbol{\sigma}][\boldsymbol{\lambda}][\boldsymbol{\psi}][\tau^2] \prod_{i=1}^{N} \{[\delta_t|\boldsymbol{\lambda}, Z_t] \prod_{i=1}^{k_t} \{[x_{t,i}, y_{t,i}|\beta_t, \boldsymbol{\sigma}, \tau^2, Z_t]\}[Z_t|\boldsymbol{\psi}, Z_{t-1}]\}
\end{aligned}
$$

# 5 Comparison Discrete vs Continuous Time models

Now that we have identified the two models in both the discrete and continuous time, [MJH+14] used Monte Carlo methods to samples from the posterior to get a Bayesian estimate of the parameters of the movement model. We will now discuss the similarities and differences in the latent states, bearing and step size in both temporal settings.

## 5.1 Latent State

Summarizing the analysis done in [MJH+14], we were able to predict the percentage of the 10 day foraging trip spent foraging, in transit, and resting along with 95% Bayesian credible regions summarized below: The breakdown of the path in given in Figure 2 in [MJH+14].

Table 1: Average time spent in each latent state

| **State** | **Discrete-time** | **Continuous-time** |
|---|---|---|
| F | 0.36 (0.26,0.39) | 0.29 (0.23,0.34) |
| T | 0.36 (0.29,0.45) | 0.61 (0.53,0.67) |
| R | 0.28 (0.22,0.37) | 0.10 (0.03,0.15) |

We also some estimates of some entries of the transition matrix $\boldsymbol{\psi}$ with their credible regions. For some reason, the authors in [MJH+14] did not list the full transition matrix, only the probability that the state doesn't change from time $t \to t+1$ and the two most likely transitions in each temporal model.

The discrete and continuous time models give very different results. The discrete-time model predicts roughly equal about of time spent in state $F, T, R$, where as the continuous-time model says that more of the time in transit versus resting. The author notes, that there were larger movement associated with resting, and small movement steps in transit, which call into question how accurate classification of the states really are. This could be be possibly be the result of the OU process being a poor choice to construct the position model.

Table 2: Transition rates

| Transition | Discrete-time | Continuous-time |
|---|---|---|
| $\hat{\psi}_{F,F}$ | 0.78 (0.67, 0.88) | 0.75 (0.62, 0.86) |
| $\hat{\psi}_{T,T}$ | 0.78 (0.65, 0.89) | 0.82 (0.75, 0.89) |
| $\hat{\psi}_{R,R}$ | 0.81 (0.71, 0.92) | 0.52 (0.10 ,0.86) |
| $\hat{\psi}_{F,T}$ | 0.15 (0.05 ,0.27) | 0.23 (0.12, 0.35) |
| $\hat{\psi}_{T,F}$ | 0.14 (0.06, 0.22) | - |
| $\hat{\psi}_{R,T}$ | - | 0.40 (0.09, 0.81) |

It can also be the poor choice of how to choose to model continuous-latent state model $Z_t$ essentially in discrete time.

The biggest notable difference in the available transition entries, is that there is a large focus on the continuous time model to enter a transit state, whereas the discrete-time process spends more time alternating between transit and foraging, both seem reasonable as a viable patter would be to forage, and once an area has been depleted, more to a new area to forage and repeat until time to rest. Again, it is difficult to discuss the transitions further without seeing the full matrix.

Although very different results, it can be argued that both models give reasonable solutions, it is also difficult to say without some expert knowledge about seal movement, which one performs better.

## 5.2 Bearing & Step Size

Although, we used very different methods to characterize the movement in the discrete and continuous time models, we can use the properties of the OU process to derive the distribution of the step size and bearing angle for the continuous time process between time $t \to t+1$. We will do this by using the result as shown in [KMT11].

**Lemma 5.1.** *Let $A \sim \mathcal{N}(\mu_A, \sigma^2)$ and $B \sim \mathcal{N}(\mu_B, \sigma^2)$ are independent.*

1. *If $L = \sqrt{A^2 + B^2}$, and $\mu = \sqrt{\mu_A^2 + \mu_B^2}$, then $[L] = R(\mu, \sigma^2)$, where $R$ is called the Rice distribution, with density,*

$$[L] = \frac{L}{\sigma^2} \exp\left(-\frac{L^2 + \mu^2}{\sigma^2}\right) I_0\left(\frac{L\mu}{\sigma}\right),$$

   *where $I_0$ is the modified Bessel function of the first kind of order 0.*

2. *If $\theta = \arctan\left(\frac{B}{A}\right)$, then $[\theta | L = l] = VM(\omega, \kappa)$, called the Von Mises. In particular,*

$$[\theta | L = l] = VM(\omega, \kappa) = \frac{\exp(\kappa \cos(\theta - \omega))}{2\pi I_0(\kappa)},$$

   *where $\kappa = \frac{l\mu}{\sigma^2}$ and $\omega = \arctan\left(\frac{\mu_B}{\mu_A}\right)$.*

Note the modified Bessel function $I_0(x)$ is given to be the exponentially growing solutions to Bessel's equation,

$$x^2y'' + xy' - x^2y = 0,$$

and can be solved using the power series and the method of Frobenius. This is a common task in an introductory differential equations class. Bessel's functions are as ubiquitous in applied mathematics as the error function is in statistics. It arrives naturally in our context as it $I_0(x)$ can be expressed as,

$$I_0(x) = \frac{1}{\pi}\int_0^{\pi} e^{x\cos\theta}d\theta.$$

Now suppose $\mu^{cts}, V^{cts}$, are the position and velocity of the continuous time movement model, we can use (11), and (13) to derive the distribution of the step-size and bearing angle denoted from time $t$ to $t+1$ by $S_t^{cts}$, and $\phi_t^{cts}$. If $l_t = \|V_t^{cts}\|$ and $\theta_t = \arctan\left(\frac{V_{y,t}}{V_{x,t}}\right)$, then

$$[S_t^{cts}|l_t, Z_t = z] = R\left(\frac{l_t(1-e^{\beta_z})}{\beta_z}, \frac{\sigma_z^2}{2\beta_z}(1-e^{-2\beta_z})\right),$$

and,

$$[\phi_t^{cts}|S_t^{cts}, l_t, Z_t = z] = VM\left(\theta_t, \frac{S_t l_t e^{-\beta_z}}{\frac{\sigma_z^2}{2\beta_z}(1-e^{-2\beta_z})}\right).$$

We can now compare the bearing angle and step size in the discrete and continuous-time setting for the different states from Figure 3 in [MJH+14]. Note that both models give roughly similar movement distributions while foraging, but drastically different results for resting and transit. The continuous-time model take significantly smaller steps than the discrete model when in transit, and significantly larger time steps when resting. This suggests the continuous time model is not robust enough to differentiate between the different states. For future efforts, having a strong criterion to differentiate the different states other that just the monotonicity condition imposed on $\beta_R > \beta_F > \beta_T$ and $\sigma_R < \sigma_F < \sigma_T$ might make the model more robust.

The continuous-time model is however significantly better at predicting the movement path. The posterior prediction for the measurement error in the discrete model was $\hat{\sigma}_x = 472m$ and $\hat{\sigma}_y = 489m$ with 95% credible interval given by (360, 596) and (381, 617) respectively compared to the measurement error in the continuous time model of $\hat{\tau} = 64m$ with credible interval (55,75). Which primarily because we did not have to do any linear interpolations when designing the measurement model in continuous time.

# 6 Discussion

## 6.1 Continuous versus discrete time models

The Northern fur seal tracking example shows that it is not always clear whether or not one show use discrete or continuous time models for mammal tracking. They both have their advantages as well as difficulties. In general it really depends on the time scales are natural to the process we want to observe, and the data we have access to.

Discrete-time models are the natural choice when your observation occur in a temporal homogeneous way and there is a natural time scale for your process and data. If the process you are observing occurs in regular time intervals then discrete time methods can be very effective. They have the benefit of being easy to define and are often easier to implement, both mathematically and computationally. Often one arrives at the continuous time models by taking scaling limits of discrete ones.

Continuous time-models are often derived by taking scaling limits of discrete-time models that we know and love. This often results in the continuous time model smoothing out the "messiness" of the discrete-time system and allow us to extract the essential properties. These scaling limits are themselves scale invariant, making them very robust to temporally irregular data, such our case with the seal data.

The canonical example to illustrate this scale invariant nature is Brownian motion and random walks. A random walk $S_n$ is a very general discrete time model characterized by the fact that $X_n \equiv S_n - S_{n-1}$ are iid with mean 0, that ubiquitous when discussing random systems. Donsker's theorem states that if $X_n$ has variance $\sigma^2$, then $\frac{S_{\lfloor Nt \rfloor}}{\sqrt{N\sigma^2}}$ converges weakly to a standard Brownian motion, independently of $X$. Thus, Brownian motion, a mathematically convenient, well-studied, continuous-time model can be used to gain insight about a potentially very complicated discrete-time random walk.

In our case, we used an OU process, as it is the scaling limit of a complicated $AR(1)$ process, is natural when discussing systems that we wish to have mean reverting behaviour.

Once a continuous-time model is defined, they have the advantage that they are very powerful in terms of insights they can give about our system. Moreover, and there has been a great deal of theory developed to classify and manipulate them. The disadvantage is that they can be very difficult to formulate and require a bit of finesse and mathematical maturity to interpret.

In defence of discrete-time systems, the fact that they are easy to design and interpret makes them a very powerful tool in the statisticians arsenal, and can be quite robust to a variety of data. They do require a lot of choices to be be made by the designer and thus designing a good model is very much an art. They can also be very useful none the less.

## 6.2 Future extensions

For future work, it is possible to design discrete time systems with irregular time steps. For example when taking a step in the measurement model between time $t_i$ to $t_{i+1}$, one can scale the step size proportional to the time window. That can eliminate the need for linear interpolation in the measurement model and can potentially improve the accuracy of the movement model.

We also did not incorporate any environment factors. It seems reasonable to assume that the animals are in taking in the environment factors to decide where to move. Seals might avoid certain areas if there are currents that are difficult to navigate, and might be drawn certain known sources of food. One active area of research in propobabilty theory is the study of random walks in random environments. It might be reasonable to model the ocean as a random environment and model the seals as performing a random walk, as opposed to just modelling the seal movement as a correlated random walk. One area of potential research might be the study of a correlated random walk in a random environment.

This seem more difficult to incorporate for tracking sea mammals, but it definitely can be incorporated in for tracking land mammals such as bears. Topographical and satellite data can be used to identify how easy certain areas of forests are to have resources that would be of interest to a bear, and to identify the difficulty of a bear to navigate to. One could incorporate into the model certain regions that would bias a bear, as opposed to just assuming purely a random walk.

# 7 References


[fas17] Wildlife computers, https://wildlifecomputers.com/data-products/fastloc/, March 2017.

[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(5):1208–1215, 2008.

[KMT11] Hisashi Kobayashi, Brian L Mark, and William Turin. *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press, 2011.

[MEP01] N Hastings M Evans and B Peacock. Statistical distributions, third edition. *Measurement Science and Technology*, Ch 41:189–191, 2001.

[MJH+14] Brett T McClintock, Devin S Johnson, Mevin B Hooten, Jay M Ver Hoef, and Juan M Morales. When to be discrete: the importance of time formulation in understanding animal movement. *Movement Ecology*, 2(1):21, 2014.

[MKT+12] Brett T McClintock, Ruth King, Len Thomas, Jason Matthiopoulos, Bernie J McConnell, and Juan M Morales. A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, 82(3):335–349, 2012.