Variation in distance between consecutive animal locations

This should be the abstract

# Introduction

Movement is a fundamental ecological process, critical to an animal’s biology and it’s interactions with the environment. The study of animal movement tends to focus on four basic mechanistic questions: the why, when, where, and how an animal moves, and these questions can be answered at the scale of individuals or populations. A general framework for movement ecology is needed to link movement patterns and processes. Such a framework would start with a focus on movement itself and the individual’s internal state, navigation capacity, motion capacity, and responses to external factors (Nathan *et al.* 2008). The overall goal of studying movement is to find the linkages between the causes and mechanisms of animal movement, that explain spatiotemporal patterns that then feed back into various ecological and evolutionary processes. However, we can also focus on the effects and consequences that different movement patterns can have over species interactions and other ecological processes.

Random walks have been used broadly in biology to model movement of animals and microorganisms, starting with the simplest models of movement using uncorrelated and unbiased random walks (Codling *et al.* 2008). These movement models show no preference in direction and movement at each step is completely independent of the direction taken in previous movements. Extensions of simple random walk processes are regularly used to accommodate for more complex behaviors or to include drift. Some of these extensions involve the use of correlated random walks, which incorporate a correlation between directions in successive movements, biased random walks, which incorporate a directional bias, or even building movement models as mixtures of random walks to account for more complex behaviors (Morales *et al.* 2004). Modeling approaches using random walks generally assume a homogeneous environment and behaviors that don’t drastically change the movement pattern (Kareiva & Shigesada 1983). A simple random walk approximates displacements by connecting a series of straight lines, and can summarize animal movement based on parameters for movement length and turning angle, or the use of net squared displacement. Connecting theoretical movement models based on correlated random walks to data had often relied on the use of these parameters or displacement measures (Turchin 1998), but more current and efficient approaches have used likelihood-based techniques to estimate parameters by fitting a discrete approximation of the model directly to the location data (Jonsen *et al.* 2003).

In the context of species interactions and the consequences of animal movement to dispersal, the general focus has been on understanding how the typical movement patterns, built around average measures at population levels, can influence plant population spread (Levey *et al.* 2005), seed dispersal (Jones *et al.* 2017), pollen dispersal (Garcia *et al.* 2007), and even pathogens (Shaw *et al.* 2017). Animal movement can be modeled with a simple diffusion process or a random walk (Levey *et al.* 2005; Jones *et al.* 2017), with the simplest approach assuming model parameters are the same across individuals, something known as ‘complete pooling’ (Langrock *et al.* 2012). This focus on generality has driven to a lack of frameworks for understanding how variation in animal movement is maintained, and the consequences of this variation across populations and communities. From the animal movement perspective, some alternatives have been considered to incorporate individual variation with ‘no pooling’ or ‘partial pooling’ approaches, where each or some individuals have their own set of parameters (Patterson *et al.* 2009), however a high number of parameters quickly limits applications to high levels of individual variation. The decision to pool data or not may also come from the modeling framework chosen, as frameworks that allow for random effects may have a better structure to incorporate this variation than simpler frameworks. Recent calls for incorporating individual variation in movement and their consequences to seed dispersal or even pathogen spread (Snell *et al.* 2019) have pushed for the need to develop theory that explores different types of variation in movement patterns to understand its consequences across ecological organizational levels. In particular, when considering species interactions, incorporating individual variation becomes essential to understand how positive or negative feedback loops play into the maintenance of variation in movement as individuals respond to external factors in their environment.

We focus here on developing a process to analyze animal movement sequences based on movement lengths, or step lengths, and turning angle probability distributions. Following the methods of Kareiva and Shigesada (1983), we assume the movement is a correlated random walk and derive expected square displacements associated to a number of consecutive moves under various probability distribution models. Unlike other studies which focus on deciphering the underlying behaviors from animal movement data, our focus was to uncover potential differences between individuals and how these differences can be obscured when modeling population-level animal movement. We know these simple correlated random walks, or simple movement models might not be enough to describe net displacements under more complex processes, for which developments in higher order markov processes show to be promising. However, our goal in this study was to use a simpler model and focus on potential differences that arise when comparing data from multiple individuals under frameworks that completely pool data in contrast to those that don’t.

# Methods

## Animal movement model

## Description of the models considered

## Specific case with telemetry data

We focused our study of individual variation on the many-banded araçari (*Pterglossus pluricinctus*), a small toucan in Yasuní National Park, Ecuador. We used previously collected data from studies using radiotelemetry to estimate home ranges and evaluate the potential seed dispersal distances for the *Virola flexuosa* tree (Holbrook 2011). Over a period of four years, from 2001 to 2005, *P. pluricinctus* individuals were captured and radio-transmitters were attached at the base of the central tail feathers (Holbrook 2011). Toucan locations were estimated by triangulation using receivers and hand-held antennas, with tracking periods lasting between 4-5 hours per day, locating individual birds every 15 minutes. Further details on field methods can be found in Holbrook (2011). It is important to note that these methods did not track all birds simultaneously, and that although individual toucan locations were attempted every 15 minutes, this was not always possible, thus some successive locations have time categories associated to multiples of 15. Due to the nature of the data, and the irregular time lags between two consecutive locations, previous work calculated movement rate for each individual in order to take advantage of all the data despite variable time intervals (see Table 1 in Holbrook 2011). Movement rates were calculated as the displacement in meters over the total length of the tracking period in minutes, where the tracking period was defined as the total active tracking time for a bird over the season when the data was collected. In other words, the movement rate could be defined as the average velocity of the animal over the entire tracking period, and thus there is one movement rate value associated to each individual bird. In our case, we wanted to extend the use of the telemetry data and use all locations in a modeling framework. From the relocation data, we calculated animal trajectories by considering discrete steps between successive relocations of each animal (Turchin 1998), quantifying their turning angle after each movement, estimating the Euclidean distance between successive relocations, and estimating the velocity for each segment as the displacement distance divided by the length of the time interval. We used the package ‘adehabitatLT’(Calenge 2006) in R (R Core Team 2020) to calculate other descriptive parameters for each trajectory, such as turning angles and net squared displacement.

# Results

The majority of trajectories obtained were regular, with 15 minute time lags between successive relocations, however, some relocations were gathered at different time intervals, making them irregular. To analyze all trajectories and compare their descriptive parameters, we decided to work under a velocity-based framework.

For our analyses, we focused on toucans with a minimum of **20?** recorded locations, and we calculated turning angles (in radians), and the associated Euclidean step lengths in meters for each successive set of locations and turning directions. We worked under a velocity-based framework, and thus divided step lengths by the time interval during which they were recorded. This gave us a total of **600ish?** data points for turning angles and velocities.red

What do we consider a trajectory? The path followed by consecutive relocations during a tracking session. The tracking session for this data is the 4-5 hour period each day that they follow the bird. Then, the number of relocations for each trajectory varies from one to 26, with a mean of 4 relocations. This is a limitation of the data, since it’s not easy to characterize a full trajectory of movement with such a small number of relocations in a trajectory.

Dealing with the issue of autocorrelation in this case: what does that mean? Or we just assume each step is independent from the others. Think I saw this in the Kareiva Shigesada paper.

# For this to run I need to move the load of data and such higher up  
aracari\_df %>%  
 group\_by(id, burst) %>%   
 tally() -> relocs   
  
# Having a minimum number of relocations for "time series analysis" or something, fitting.  
relocs %>%   
 filter(n >= 5) %>%   
 distinct(id)  
  
  
relocs %>%   
 ggplot(., aes(x = n, fill = id)) +  
 geom\_histogram()  
summary(relocs$n)

aracari\_df %>%   
 filter(R2n >0) %>%   
 ggplot(., aes(x = R2n)) +  
 geom\_histogram() +  
 facet\_wrap(~id)

**Could I find anything pre HMMs of SSMs? Wondering what people could do before? And wondering this because we have broken up tracking periods, not continuous ones**  
Seems like I need to focus more on random walks

Assume weibull distributions for step length and wrapped cauchy distributions for the turning angles (Langrock *et al.* 2012), but they tried Gamma and von Mises, respectively, but got outperformed by AIC. Used negative binomial for the state dwell times.

Random walks used broadly in biology to characterize animal and cell movement. Describes que commonly used distributions for angles: von Mises, wrapped normal and wrapped Cauchy distributions. (Codling *et al.* 2008) - IDEA: so maybe use these distributions for the angles. We might see that there is not much difference in angle movement across individuals? They all tend to move forward perhaps, or something like that, and where the real differences are is in the distances. - Also might be intriguing to incorporate the time as a set of multiple random walks? or repeats the velocity value for that number of minutes. Maybe just using a correlated random walk, which takes into account short-term correlations in the direction of movement? - Codling describes this as a velocity jump process, since the markov process is the walker’s velocity rather than the location. Details at the end of page 7. - We consider the individuals moving in space at a constant speed , and at each time step each individual changes direction and moves a distance in a new direction (with probability ), or moves a distance in the previous direction (with probability ). Hence, turning events occur as a Poisson process with rate . **This is textual from paper, so modify** Their focus is also on multiple individuals at the same time, and the density of individuals at a given time and location. - In my case, the distance would also be randomly sampled, or this are the velocities I suppose. So, perhaps it is constant speed, and the distance is actually times to account for 15, 30, etc minutes based on the data. - Remember too that this velocity is actually the distance over time - Pg11. A correlated random walk, CRW, consisting of a series of independet draws from a step length PDF and a turning angle PDF for each step, a first order markov process. - using mean square displacement. The effect of step length variability on MSD can be quite significant. - MSD being of interest to ecologists due to its relation with the diffusion coefficient D. equations 2.13 and 2.14 \*\* Should write a conclusion about this paper\*\*

(Kareiva & Shigesada 1983) - approximating displacements by connecting a series of straight lines, and thus summarize animal movement based on parameters for movement length and turning angle. Using square displacement instead of linear displacement, because it’s expectation can be calculated from turning angles and step lengths. Assume that the length of each move and the turning angle are independent random variables, each with its own probability density. Assume that a series of moves can be represented by random draws from these probability densities, and since each random draw is independent from the preceding ones, the random draw process is a first order markov chain. The result is a correlated random walk, and the distribution g(theta) gives a measure of the degree to which the angles of movement are correlated. *Would be interesting to add a figure with what I am actually analyzing. So, visually show how the gps locations get turned into straight lines, and angles, with an associated time interval. Or we could just multiply everything? So have it all for 15 minute intervals, and then repeat the ones that went on longer, perhaps set a limit for 60 minutes?*

probabilistic rules of movement.

As in other papers. Note that we adopt the convention of using upper case letters for random variables and lower case letters for possible numerical values of these random variables.

“Location accuracy using radio telemetry may be reduced in tropical forests… conservative distance categories of 100-m increments to better represent precision. Time categories: 15, 30, 60, 90 minutes. Then calculated probability of movements made within each distance category within each time category, summed across each time category to give a final probability for each distance category.”

“Distances travelled per movement bout ranged from 0 to >2000m. Strong leptokurtic distributions of movements with most being <300m. Longest recorded movement for Pluricinctus is 3665m (recorded in 30 minutes)” \*this is a corresponding velocity of 122.1666667 meters per minute.

Evaluating models with their individual-specific counterparts, using AIC or BIC. In individual-specific models the AIC results from the joint likelihood of the individual-specific models.

Consider incorporating NAs for missed observations so we keep the regular 15 minute time interval.

### MPM: meters per minute

Due to the variation in time between consecutive locations, we scaled the step lengths to displacement in one minute. **Visualize the variation and distribution of these distances between locations**

# Fit distribution

## Information criteria

However, it is not as straight-forward since there are multiple competing models for certain individuals, which we consider as those within less than 2 units of the delta AIC or BIC. When using AIC, there are competing models for almost every individual, whereas with BIC, only 4 of the individuals have competing models. We should use 4 for BIC and 3 for AIC. Look for the paper Taper Ponciano 2016.

Now we can compare what we call *models*, in which we compare a model that does complete pooling and considers all the data together, and the other model that considers individual variation, and so it fits a separate distribution to the data from each individual.

The individual variation models will fit a distribution to each individual. At one level we will use the same distribution for all individuals and only consider variation in parameters. At the next level, we will consider a change in parameters and distributions, where individuals can have different distributions, and this is what we call the mixed distribution model for individual variation.

Considering just focusing on BIC because of JMP’s paper, and also in this specific case, it provides less competing models for each individual

## Probability distribution parameters

We can compare the different parameter values estimated for each distribution and their standard deviations. These parameters can give us some insight into some of the characteristics of each set of distances moved per minute, such as which individuals have higher means or longer tails in the distribution of their distances moved per minute. The x axis in the following plots highlights which individuals have the lowest BIC (yellow) or AIC(magenta) with that model. The blue lines show the value of that parameter for the population level model, and dashed blue lines show the upper and lower limits of that parameter estimate. Overall, at the population level, the lognormal model had the lowest AIC and BIC values.

#References

Calenge, C. (2006). The package adehabitat for the r software: Tool for the analysis of space and habitat use by animals. *Ecological Modelling*, 197, 1035.

Codling, E.A., Plank, M.J. & Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal society interface*, 5, 813–834.

Garcia, C., Jordano, P. & Godoy, J.A. (2007). Contemporary pollen and seed dispersal in a prunus mahaleb population: Patterns in distance and direction. *Molecular Ecology*, 16, 1947–1955.

Holbrook, K., Loiselle, B. & others. (2007). Using toucan-generated dispersal models to estimate seed dispersal in amazonian ecuador. *Seed dispersal: theory and its application in a changing world*, 300–321.

Holbrook, K.M. (2011). Home range and movement patterns of toucans: Implications for seed dispersal. *Biotropica*, 43, 357–364.

Jones, L.R., Duke-Sylvester, S.M., Leberg, P.L. & Johnson, D.M. (2017). Closing the gaps for animal seed dispersal: Separating the effects of habitat loss on dispersal distances and seed aggregation. *Ecology and evolution*, 7, 5410–5425.

Jonsen, I.D., Myers, R.A. & Flemming, J.M. (2003). Meta-analysis of animal movement using state-space models. *Ecology*, 84, 3055–3063.

Jonsen, I.D., Myers, R.A. & James, M.C. (2006). Robust hierarchical state-space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology*, 75, 1046–1057.

Kareiva, P. & Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologia*, 56, 234–238.

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

Levey, D.J., Bolker, B.M., Tewksbury, J.J., Sargent, S. & Haddad, N.M. (2005). Effects of landscape corridors on seed dispersal by birds. *Science*, 309, 146–148.

Morales, J.M., Haydon, D.T., Frair, J., Holsinger, K.E. & Fryxell, J.M. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, 85, 2436–2445.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R. & Saltz, D. *et al.* (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences*, 105, 19052–19059.

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

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shaw, A.K. (2020). Causes and consequences of individual variation in animal movement. *Movement ecology*, 8, 1–12.

Shaw, A.K., Peace, A., Power, A.G. & Bosque-Pérez, N.A. (2017). Vector population growth and condition-dependent movement drive the spread of plant pathogens. *Ecology*, 98, 2145–2157.

Snell, R.S., Beckman, N.G., Fricke, E., Loiselle, B.A., Carvalho, C.S. & Jones, L.R. *et al.* (2019). Consequences of intraspecific variation in seed dispersal for plant demography, communities, evolution and global change. *AoB Plants*, 11, plz016.

Turchin, P. (1998). *Quantitative analysis of movement: Measuring and modeling population redistribution in animals and plants*. Sinauer Associates.