

CHAPTER 3

VELOCITY-BASED ANALYSIS OF RADIO TELEMETRY DATA TO EXPLORE INDIVIDUAL VARIATION IN ANIMAL MOVEMENT

Movement is a fundamental ecological process, critical to an animal's biology and its interactions with the environment. Animal movement, in particular, plays an important role in the structure and dynamics of plant populations through the effects of frugivory and seed dispersal services (Côtés and Uriarte 2013). From the seed dispersal literature, a common approach to understand how animal movement contributes to seed depositions is to generate seed shadows, or the spatial distribution of seeds for a single source plant (Nathan and Muller-Landau 2000). Generally, seed shadows are generated by combining gut retention times, the amount of time a seed remains ingested by an animal, with animal movement data and environmental covariates to understand spatial seed deposition patterns. Usually involving an individual-based modeling framework, plant, animal, and environmental information is then used to estimate the effects of plant density or habitat loss and fragmentation on spatial seed deposition patterns (Levey et al. 2005, Morales and Carlo 2006, Lenz et al. 2011, Jones et al. 2017, Pegman et al. 2017). Other studies have focused on incorporating different animal behaviors (Russo and Augspurger 2004, Russo et al. 2006, Karubian and Durães 2009, Sasal and Morales 2013, Bialozyt et al. 2014) and more recently, attention has been brought to the importance of variability in seed disperser assemblages (Jordano et al. 2007, Rehm et al. 2018, Zwolak 2018, Snell et al. 2019). Despite the complexity of animal movement, models used to estimate seed shadows make significant simplifications of the movement processes as the focus is on the dispersal of seeds (Côtés and Uriarte 2013). These models also generalize across populations, likely overlooking the important consequences that individual variation in

seed-dispersing animals may have for plant populations and communities (Zwolak 2018).

The scientific study of animal movement has received significant attention over the last decade, with great advances in both the technology to track movement, and the quantitative methods used to analyze movement (Nathan 2008, Joo et al. 2020). However, the focus has been on trying to analyze how the environment affects movement and behavior (Joo et al. 2020), which has led to the development of complex quantitative techniques, in particular state-space modeling frameworks. In movement data, a state-space process allows us to couple a statistical model for the observation method, accounting for errors in sampling or detection, with a separate model for the movement dynamics which are determined by the effects of an animal's behavior or responses to the environment (Patterson et al. 2008, Langrock et al. 2012, Auger-Méthé et al. 2021). However, from the seed dispersal ecology perspective, the important question is how does animal movement influence seed dispersal and thus create changes to plant communities? Despite the quantitative progress in the field of movement ecology, it has not been translated to seed dispersal frameworks, in which animal movement is often modeled as a random walk (Levey et al. 2008, Jones et al. 2017) or a constant movement rate (Carlo and Morales 2008), ignoring animal behavior or individual variability.

Although movement ecology has developed the tools to analyze complex movement, the data requirements do not always align with the movement data collected for seed dispersal studies. Movement data collected for studies of seed distribution is often collected at scales that correspond to gut passage times (Westcott et al. 2005,

Holbrook and Loiselle 2007, Holbrook 2011, Rehm et al. 2019, Nield et al. 2020), and so we lack the finer scale resolution necessary to implement more complex movement models. Additionally, data collected with radio telemetry devices, especially older ones, is subject to environmental variables such as heavy cloud cover, or dense forests, which can result in data collection at irregular time intervals, thus complicating the application of discrete-time movement models (Jonsen et al. 2003, 2005), and force the researcher to rely on interpolation, subsampling, or aggregation of data during pre-processing. To overcome this, Johnson et al. (2008) proposed using a continuous-time velocity process capable of handling irregular time intervals, and incorporated it under a state-space framework to estimate parameters and locations. However, their model assumes a normal movement process, under which heavy tailed processes (such as flight behaviors mixed with smaller movements) are more challenging to handle. Additionally, several of the movement models previously mentioned rely on some variation of a correlated random walk, under which animal locations are dependent on all previous locations, and thus require good time series data and long enough animal trajectories with multiple consecutive locations.

Despite the advances in modeling approaches used in movement ecology, they still typically focus on understanding the mean or modal movement patterns and ignore the fact that movement patterns may be widely variable even across individuals of the same species. Thus these models still do not consider the potentially complex population-level effects of individual variation (Spiegel et al. 2017, Zwolak and Sih 2020, Shaw 2020). Not surprisingly, the same is true in seed dispersal ecology, where individual variation in frugivore behavior or disperser movement patterns have only

recently started to be explored (Schupp et al. 2017, Zwolak 2018, Snell et al. 2019). Movement ecologists have started to tackle this issue by evaluating the effects of varying levels of data pooling, where 'complete pooling' is the simplest approach and assumes that all individuals are identical and represented by a single set of model parameters (Langrock et al. 2012) and thus all individual are included in estimates of those parameters, but a 'no pooling' or 'partial pooling' approach, assumes each or some individuals have their own set of parameters (Jonsen et al. 2006, Patterson et al. 2009). However, assuming larger numbers of unique movement parameters quickly outstretches the limits of the data and thus the application of models with high levels of individual variation. The decision to pool data or not may 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.

Our goal in this study was to reanalyze available relocation data with a velocity-framework to overcome challenges such as irregular time intervals between relocations and varying number of observations between individuals. Additionally, to explore how individual variation in animal movement can be translated to different parameters on movement distributions, we used a data pooling approach and assessed the fit of our models. We used a velocity-based approach to analyze radiotelemetry data from twelve *Pteroglossus pluricinctus* (Aves: Ramphastidae) individuals, a major seed disperser for *Virola flexuosa* (Myristicaceae), a tropical nutmeg, and explore the underlying differences in movement across individuals. With the data pooling approach, we compared the differences in movement at a population, social group, and individual levels using complete, partial or no pooling scenarios. Our study also relies on the use

of a nonparametric bootstrapping technique to calculate percent support for best fitting models and address unequal sample sizes for individuals. We find that incorporating individual variation for movement models is highly supported by our work, with strong evidence using the nonparametric bootstrap. However, when estimating parameters, we observe a trade off between the sample size and the number of parameters being estimated, with the strength of evidence from the nonparametric bootstrap. Models with greater sample sizes and less parameters to be estimated showed better fits but had weak evidence to support them.

Methods

Empirical Velocity-Based Analysis for the Many-banded Aracari Relocation Data

We focused the empirical aspects of our study of individual variation on the many-banded araçari (*Pterglossus pluricinctus*), a small toucan in Yasuní Biosphere Reserve, Ecuador. We used previously collected data from studies using radiotelemetry from twelve *P. pluricinctus* individuals (Holbrook 2011). Toucans were tracked over sessions lasting up to 5 hours per day, with birds located every 15 minutes to the extent possible. Animal locations were not recorded automatically, they were estimated by triangulation using receivers and hand-held antennas. The choice of the 15-minute time interval was based on gut retention times for the seeds of *Virola flexuosa*, a tropical nutmeg for which *P. pluricinctus* disperses seeds. Further details on field methods can be found in Holbrook (2011).

We used the package `adehabitatLT` (Calenge 2006) in R version 4.0.2 (R Core Team 2020) to calculate animal trajectories based on location data, and examined step lengths, turning angles, and the time intervals between these observations. Given the irregular time intervals for location points (Figure 3-1), we focused on instantaneous

velocity estimation, and calculated speed between consecutive relocations by dividing the step length travelled by the time interval between relocations. We used a permutation approach to examine potential autocorrelations between velocity data points (Dray et al. 2010). This is a common approach in movement ecology given that successive observations tend to lack independence, something relevant for high sampling frequencies. Previous approaches to the data used in this paper had relied on estimating overall movement rates as the average distance moved per minute over the whole tracking period and thus reducing the number of observations to one movement rate per individual (Holbrook 2011). A tracking period includes all the daily tracking sessions for an individual, and as our goal was to understand the variation in movement between individuals, we developed a framework that allowed us to use each observation instead of their average.

We assumed the framework of a stochastic movement process, uncorrelated and unbiased for simplicity (i.e., the turning angles of each step are independent of the previous directions, and their turning angles follow a uniform distribution), where the length of each step follows a given probability distribution (Gamma, Lognormal, or Weibull in our case) thus resembling a continuous-time analog of a random walk, or a Levy walk (where the step length follows a power-law distribution). Under this framework, given that turning angles are drawn from a uniform distribution, assume independence between angles and step lengths. Therefore, we focused on the distribution of step lengths, where each observation becomes a measured velocity (in meters per minute) estimated as the absolute distance between two consecutive locations divided by the change in time following:

$$v(t) = \frac{\sqrt{(u_x(t + \Delta) - u_x(t))^2 + (u_y(t + \Delta) - u_y(t))^2}}{\Delta} \quad (2-1)$$

In this way, we are able to analyze the movement of the animal by using all available relocations, despite irregular time intervals or short time series. In Johnson et al. (2008), the authors used a bivariate process to describe velocity in both directions, and then integrated to determine an animal's location at time, t :

$$\mathbf{u}(t) = \mathbf{u}(0) + \int_0^t \mathbf{v}(\tau) d\tau \quad (2-2)$$

This integral describes the location at time t as the sum of steps determined by velocities in each direction plus a starting location at time $t = 0$. Our framework is different because of data constraints, where the time series are short and data is limited. However, under our framework, with the assumption of uniformly distributed turning angles, we can still use the distribution of velocities to estimate the location of an animal at a given time t , and so we focus on estimating the best model to describe velocity while incorporating variation among individuals.

Unlike the estimation of velocities, which require two consecutive relocations, estimating turning angles requires a minimum of three consecutive relocations, thus making the data even more limited. However, we explored the distribution of angles using a simple and commonly used distribution, the Wrapped Cauchy (Morales et al. 2004, Langrock et al. 2012). A special property of the wrapped Cauchy distribution is that as its parameter ρ approaches zero, $\rho \rightarrow 0$, the distribution becomes a uniform distribution for a circle (Hooten 2017, p.163), which aligns with our initial assumption of an uncorrelated and unbiased random walk.

Estimating Levels of Heterogeneity by Incorporating Data Pooling

In order to assess the level of heterogeneity in movement between individuals, we used a pooling approach (Patterson et al. 2009, Langrock et al. 2012), where we varied probability density and parameter values for the distribution of velocities at the population, individual, and social group levels. We did this by fitting the models to different subsets of the data and then, we proceeded to use Bayesian Information Criterion (BIC) to evaluate the relative fit of each of the models. We use BIC as a means to compare the relative fit of these models to the data, and the effect of pooling levels as the pooling approaches that incorporate individual or group heterogeneity increase the number of parameters to estimate with smaller sample sizes.

The pooling approach used to evaluate the distribution of velocities considers separating the data into subgroups and then fitting a distribution to each of the subgroups when applicable. The choice of distributions for velocity (gamma, lognormal, and Weibull) was based on commonly used distributions in the seed dispersal and animal movement literature (Levey et al. 2005, 2008, Morales and Carlo 2006, Will and Tackenberg 2008, Morales et al. 2010, Langrock et al. 2012, Nield et al. 2020). In particular, the Weibull distribution is a very flexible distribution with a long tail, which allows for rare fast movement rates, it also is equivalent to the exponential distribution when its shape parameter is equal to 1 (Hooten 2017, P. 162). The overall description of the models is as follows:

“Complete pooling”- CP: This is the simplest approach, in which we assumed that the model parameters were identical for all individuals, and thus pooled all the velocity data to estimate distribution fits and parameters. Essentially, we assume that the

probability density describing velocities, $h(\sigma)$, follows a Gamma, Lognormal, or Weibull distribution with a single set of parameters σ shared by all individuals.

- Model 1: $h(\sigma) \sim \text{Gamma}(a, r)$ where the shape (a) and rate (r) parameters are identical for all individuals and across the population.
- Model 2: $h(\sigma) \sim \text{Lognormal}(\mu, \theta)$ where the meanlog (μ) and meansd (θ) parameters are identical for all individuals and across the population.
- Model 3: $h(\sigma) \sim \text{Weibull}(a, b)$ where the shape (a) and scale (b) parameters are identical for all individuals and across the population.

For each of these models, likelihood calculation, and thus BIC calculation, is straightforward:

$$BIC_i = -2\log(L_i) + k_i \log(n) \quad (2-3)$$

Where L_i is the likelihood for each of the models: gamma, lognormal, or Weibull. The number of parameters for each model, k_i , is two for each of the models, and the number of observations, n , is the same for all since the same dataset is being used to fit the three different distributions.

“Partial pooling” – PP: We used the information available on individuals belonging to social groups, and thus grouped the data accordingly to the seven social groups identified in the field. Thus, we estimated the best fit movement models for each social group and asked whether there was significant movement variation at this level. We explored this with four models, in three of which we allowed for heterogeneity in parameter values only, and the fourth model, multi-distribution model, we relaxed the assumption that movement across groups was drawn from the same probability distribution (gamma, lognormal, or Weibull) and allowed for variation in both parameter values and probability distribution between social groups. In essence, this multi-distribution model explicitly recognizes the heterogeneity across groups in both,

parameter values and model family. This explicit modeling of the heterogeneity differs from fitting an overall random effects model, where it is assumed that both the choice of distribution (gamma, lognormal or Weibull) and of its parameter values are random. Each of the models is described as a sum of the submodels, where each submodel is fitted to the data associated to the social group:

$$h_x(\sigma) = \sum_{s=1}^7 h_x(\sigma_s) \quad (2-4)$$

Each model follows:

- Model 4: $\mathbf{h}_{\text{Gamma}}(\sigma) \sim \text{Gamma}(a, r)$ where the shape (a) and rate (r) parameters vary for each of the social groups.
- Model 5: $\mathbf{h}_{\text{Lognormal}}(\sigma) \sim \text{Lognormal}(\mu, \theta)$ where the meanlog (μ) and meansd (θ) parameters are allowed to vary for each social group.
- Model 6: $\mathbf{h}_{\text{Weibull}}(\sigma) \sim \text{Weibull}(a, b)$ where the shape (a) and scale (b) parameters vary for each social group.
- Model 7: $\mathbf{h}_{\text{multi}}(\sigma) = \sum_{s=1}^7 h_x(\sigma_s)$, where each social group is characterized by the best fitting probability density and their associated parameters.

For each of the four models, we calculated BIC as the sum of the likelihood of each submodel, minus the total number of parameters estimated, and accounted for the sample size in each of the social groups.

$$BIC_i = \sum_{s=1}^7 k_{i,s} \cdot \log \left(\sum_{s=1}^7 n_s \right) - 2 \sum_{s=1}^7 \log(L_{i,s}) \quad (2-5)$$

“No pooling” – NP: The setup of models for this scenario is similar to the partial pooling approach, except that there is no pooling of data, and we allow for heterogeneity in parameter values for each of the twelve individuals. Similarly, for models 8, 9, and 10, individuals only vary in their parameters, but the family of the distribution is gamma, lognormal, or Weibull respectively. We describe a multi-

distribution model, which allows for variation in probability distribution and parameter values across individuals.

Given the heterogeneity of sample sizes, we performed a non-parametric bootstrapping technique to assess the reliability and support for each model following suggestions in (Taper et al. 2021). Essentially, we performed sampling with replacement for each of the velocity dataset to generate 1000 data sets for each group, corresponding to their pooling level. We then estimated the likelihood and BIC for each model under each bootstrap replica and determined the percentage that each model was favored by the lowest BIC score. We compared the percent support given by the nonparametric bootstrapping to the original model fits and evaluated the variation in parameters at the social group and individual levels.

The justification for this non-parametric technique has its roots on recent theoretical and practical contributions by Dennis et al. (2019), Lele (2020a, 2020b) and Taper et al. (2021). Additionally, Cox (1958) writes that although an inferential statement about the parameters of interest is any statement regarding the form of the underlying mechanism or a future outcome, this statement becomes a statistical inferential statement only when a measure of uncertainty is attached to it. The authors above show that the estimation of the sampling distribution of the ΔBIC via non-parametric bootstrap is a much more reliable measure of uncertainty than any such measure provided by either classic Neyman-Pearson approaches or a Bayesian approach in particular when none of the models are perfect representations of reality, as it is the case here. Under this scenario, the classical Neyman-Pearson approach, which depends critically on one of the models tested being a perfectly correct representation

of reality, fixes the Type I error probability irrespectively of sample size and thus problematically assesses the evidence against the null hypothesis but remains silent with respect to the evidence for the null. The decision of picking an alternative over a null hypothesis is not controversial per se, however, the probability of erroneously choosing the alternative when the null is true remains stuck at the chosen alpha level regardless of how large the sample size is. Matters get much more complicated when it is considered that the original Neyman-Pearson theorem assumes that the data were generated under one of the two models but provides no guidance whatsoever in the case of model misspecification. Although detailing the theoretical and practical characterizations of the reliability of our non-parametric bootstrap technique is beyond the scope of this work, it is important to note that under this approach, the probability of making the wrong model choice is controlled as sample size grows large despite model misspecification. Despite knowing in advance that our statistical models are mere approximations of reality, our approach based on the BIC is guaranteed to work better than the two other main statistical paradigms: the classical Neyman-Pearson Hypothesis testing and Bayesian approaches.

Results

Exploring the Variation In Speed and Sampling

Original data collection methods attempted to relocate individual birds every 15 minutes (Holbrook and Loiselle 2007), however it wasn't always possible to obtain a relocation then and so there is considerable variation in the number of observations for each individual and the time interval between these observations (Figure 3-1). The original dataset contains 907 bird relocations, however, to estimate movement rates or speeds, we need at least two consecutive relocations, which translated to 670

Commented [JR1]: Should be a statement that summarizes the finding here.

observations for a total of 12 individual birds, ranging from 32 to 89 velocity data points. Although the majority of data points were collected at 15 and 30 minutes (55% and 30% of observations, respectively), the time intervals ranged from 15-210 minutes for consecutive relocations (a detailed breakup of the distribution of these relocations is shown in Figure 3-1). The median number of relocations for each individual was five, which converts to four velocity data points, with the minimum being two and the maximum 25. These consecutive relocations are the time series used for more complex movement ecology models, such as state-space models, and in movement ecology they are called bursts. In the case of our data, the burst of 25 relocations for B19 are an outlier, as the majority of the data is contained in bursts of 2 to 7 relocations (min and third quantiles), and the number of bursts per individual ranged from 5 to 17. To work with relative angles, which are used to determine changes in directions, a minimum of three consecutive relocations are necessary: the first two relations establish the initial trajectory working as the reference, so the turning angle is estimated as the change from the initial trajectory to the direction of the third relocation point. In the case of the angles, we had 528 data points to work with, across all 12 individuals, with a range of 25 to 70 angle data points per individual. The histograms in polar coordinates exemplify some of this variation in the number of observations (variation in the height of bars) but also show the variation in directions (Figure 3-1).

We estimated probability density curves based on the instantaneous speed data following a no pooling, partial pooling, and complete pooling approaches. By so doing, we explicitly characterized the variability across individuals and social groups of the many-banded araçari. We found considerable variation in the distribution of

instantaneous velocities when comparing the partial and no pooling approaches to the baseline of the complete pooling approach (Figure 3-2). The black line in Figure 3-2 shows the density of velocities generated by the CP approach, which pools all the data together and assumes individuals are identical to each other. This CP density dampens any variations at the level of individuals and social groups, which is very clear at the tail of these distributions, as the CP density line (black) has a significantly thinner tail. Although all scenarios presented leptokurtic distributions of speed, some individuals and social groups present higher peaks at 120, and 150 meters/min compared to the complete pooling density curves.

Distribution Model Fits Across Pooling Levels

Under the complete pooling scenario, we found that the lognormal distribution fitted speed data the best, relative to the gamma and Weibull models. However, in both the partial and no pooling scenarios, the multi-distribution model was the best fitting one when comparing within the same pooling category (Table 3-1). However, when comparing all models across the three pooling scenarios, the multi-distribution model under a partial pooling scenario had the lowest BIC score of all models, followed by the lognormal model in partial pooling, and the lognormal under the complete pooling scenario. Models incorporating individual variation, the NP scenarios, were the worst fitting models, falling behind all PP and CP variations.

Given unequal sample sizes between the three pooling scenarios, and between individuals and social groups, we used a nonparametric bootstrap approach (Taper et al. 2021) to estimate parameters under the three different distributions for each pooling scenario. In the case of partial and no pooling, parameters were estimated for each submodel for every bootstrap, and then using BIC, the best model was selected for that

bootstrap. We found that the best distribution for each group varies, and that parameter estimates also differed considerably (Figure 3-3).

Results from the nonparametric bootstrapping technique showed that the best sub model selected for individuals or social groups with all the data would generally align with the highest percentage of support (Figure 3-4). However, the percent support for each model varies greatly across individuals and social groups, with some models receiving close to 100% support and aligning with the initial best model from the full data run, whereas others received close to 30% support for each distribution (such as individual B84) showing that any of the three distributions could likely be used to represent the speed for that individual or group. At the larger level, we observed that the complete pooling model received high percentage of support for the lognormal distribution, something that aligns with the initial distribution fitting. For both the partial pooling and no pooling scenarios, the multi-distribution model received 100% support. We visualized the distances in BIC units from each model to the best model for all bootstrap replicas (Figure 3-5), showing that in general, the gamma distribution tends to be the worst model for all pooling scenarios. The distance between the best model and the second-best model is considerable for all three pooling scenarios (Table 3-2), with an average of 15 BIC units between them. However, with particular focus on the lower quantile (0.025), we observe a difference between the NP and PP scenarios and CP, which is also evident in Figure 5, where the lower quantile for CP is less than 1, providing little or no evidence to support the best model compared to the second best model. In the case of PP and NP, the lower quantile has a values close to 7 and 9, providing strong evidence for selecting the multi distribution model as the best choice.

Wrapped Cauchy Distribution For Angles

The results from fitting a wrapped Cauchy distribution (Figure 3-6) to turning angles shows variation in parameter estimates across individuals and social groups, with the estimates using all data under a complete pooling scenario aligning with the median values of parameter estimates for the other two scenarios. Previous studies have focused on categorizing movement based on step lengths and turning angles, determining encamped or exploratory positions (Morales et al. 2004). The simplest random walk assumes a uniform distribution of angles, whereas a correlated random walks (CRW) occurs when turning angles are concentrated around zero (Turchin 1998). In this case, we used a wrapped Cauchy distribution, for which the μ parameter determines the mean direction, and the ρ the mean cosine of the angular distribution (Morales et al. 2004). We observe variation in the distribution of turning angles (Figure 3-1), which shows also in variation in parameter estimates, however, further analysis of changes in directionality to determine type of behavior are beyond the scope of this paper.

Discussion

The main result of this work is that individual variation in movement behavior very likely shapes the rarity of long-dispersal events and as such, may play a key role in the structure and organization of tropical communities. Ignoring this individual variation may lead to a severe underestimation of the chances with which rare dispersal events may occur. These results have wide implications not only for animal movement studies, but for general ecological studies of community assembly in complex ecosystems like this Ecuadorian tropical forest.

Commented [BL2]: This conclusion seems like a big jump to me as a reader and I'd want to know where it comes from. I think it comes from Figure 2 and the longer fat tails – but this is velocity, so one still needs to say that moving faster assumes that this results in more LDD... My point here – just remind us where this conclusion comes from explicitly. (You do explain this in the next paragraph... so probably OK as is).

Commented [BL3]: Some might question what you mean here... can you give a parenthetical example or two; and is this just associated with complex systems, or would it be also true in say temperate forests?

Arguably, it is important for maintenance of tree tropical diversity, say perhaps through rare dispersal advantage, genetic neighborhood sizes, ...

One of the key assumptions of classical population dynamics models is that populations are large and well mixed, which leads to many of these models representing a ‘mean field’ perspective of population or community dynamics. However, over the last decade, a growing interest on variation (Violle et al. 2012, Holyoak and Wetzel 2020) and the ecological consequences of this variation calls for a shift from ‘mean field’ approaches to frameworks that integrate individual variation. From both the animal movement and seed dispersal perspective, evidence suggests that individuals can experience their environment very differently and have responses that are vastly different from the average (Matthiopoulos et al. 2015, Zwolak 2018, Snell et al. 2019, Schupp et al. 2019). For example, (González-Varo and Traveset 2016) found that the thresholds determining ‘forbidden’ species interactions are easily changed when including intraspecific variability, having direct consequences over network connectance. In our specific scenario, the variability found in the distribution of velocities, particularly the variation in the tail end of the distribution, has implications for estimating how far and how fast animals may move, and move seeds with them. We observe how an approach that ignores differences between individuals dampens the variation of the tail distribution of velocities (Figure 2 insert plots, comparison to the CP black line). Additionally, an approach that ‘pools’ individuals under the same category limits the parameter region occupied, and as we observe in Figure 3 the variation in parameter space occupied by different individuals and social groups is much wider than that of the complete pooling scenario. Previous work has shown that reducing intraspecific variability to a single average can lead to underestimating connectedness in biological networks (Poisot et al. 2015), but our understanding of how this variability

changes seed dispersal estimates is extremely limited (Snell et al. 2019). Previous work by Rudolph et.al. (2022) has demonstrated, from a simulation perspective, that variation in animal movement rates translates to seed dispersal kernels, and models incorporating individual variation predict a longer and more frequent long-distance dispersal events. The animal movement literature has found similar evidence on the importance of heterogeneity in individual-level movement for uncovering behaviors and scaling to population level patterns (Skalski and Gilliam 2000, Morales et al. 2004, Mueller and Fagan 2008, Mueller et al. 2011).

Velocity Framework Is Useful to Deal With Sparse Data

The increased interest in individual variability is also tied to the advances in tracking technology and the large influx of complex and detailed data, making the application of new and more complex statistical approaches a reality (Langrock et al. 2012, Patterson et al. 2017). The novel statistical approaches developed for animal movement are based on time series analysis requiring long series of consecutive relocations, a common characteristic of electronic tracking devices. However, the radiotelemetry data used in our study was collected in the early 2000s, with handheld antennas, as Kimberly Holbrook and her team followed individual birds for 4-5 hours a day in dense rainforest. As we presented in the results, the number of consecutive relocations for each animal in this dataset is variable, and so is the number of bursts of consecutive relocations and the time interval between relocations. The approaches used in the past involve calculating movement rates, a single value for each individual estimating the average displacement per time unit that was observed during the whole tracking season (Holbrook 2011). Using a velocity-framework allowed us to take advantage of the available data despite uneven time intervals, to estimate distribution

functions that describe the velocities observed, and incorporate the variability across individuals and social groups.

Incorporating Individual Heterogeneity

Although use of pooling approaches has been suggested in previous research they have been done at highly complex levels with state space modeling frameworks that allow incorporation of random effects associated to the variation between individuals (Patterson et al. 2009, Langrock et al. 2012). However due to the data constraints previously mentioned, using a very complex state-space model is not possible with our data. Additionally, our model differs from fitting an overall random effects model in that it explicitly recognizes the heterogeneity across groups not only in parameter values, but also model family.

To overcome the variability in observations per individuals we use this nonparametric bootstrapping approach to calculate percent support for each of the models. It's interesting to see that their own individuals are firmly supported with the bootstrap UM to a single distribution model whereas others exhibit more variation and. There are many reasons why this variation could happen ranging from the home range size nesting status nest roosting location uhm and even the dynamics between social groups. Animal personalities comes into place well as more bold individuals can probably go farther go faster other places.

Considerations For Future Research

Seed dispersal is very complex and context dependent, it requires the integration of various fields of research and both theoretical and empirical approaches (Nathan and Muller-Landau 2000, Beckman et al. 2020). Recent efforts have been made to bring together the perspectives of seed dispersal and movement ecology (Borah and

Commented [BL4]: ?

Commented [BL5]: Lost you in this sentence.

Commented [BL6]: From Borger et al: "Home ranges are spatial expression of behaviors animals perform to survive and reproduce (Burt 1943)". Individual movements are influenced by "individual state" and external environment. To take these ideas, then individual variation in movements might have been influenced by physiological conditions or behavior related to reproductive status (non-breeding, incubating or feeding fledglings), distribution of feeding trees within home range which varies spatio-temporally, predation risk, social context, among other.

Borger et al. 2008 Ecology Letters 11:637-650
Burt 1943 Journal of Mammalogy 24:346-352.
Shaw 2020 Movement Ecology 8:
<https://doi.org/10.1186/s40462-020-0197-x>

Beckman 2021) to advance the mechanistic understanding of animal-mediated seed dispersal. Although the current technology to track animal movement allows for high temporal resolution in the data, several seed dispersal studies performed in the last two decades have valuable radiotelemetry data that are sparse and have shorter time series. The approach we developed in this study is a step towards incorporating complexity and individual variation despite the lack of high-resolution data. We have shown that individual variation in animal movement can be detected with this approach, and that multi distribution models outperform single distribution models, despite estimating a higher number of parameters from smaller data frames

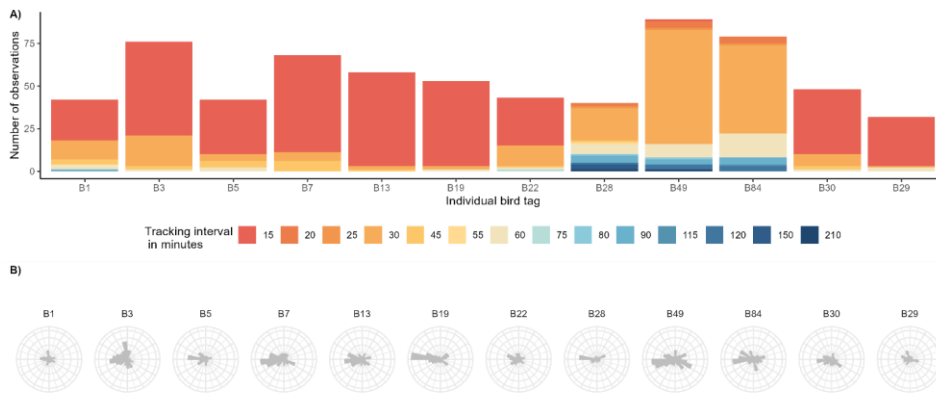


Figure 3-1. A) Breakdown of the time intervals in the available data for each individual tagged bird. The majority of relocation points are recorded at 15 or 30 minute intervals. However, the intervals between relocations can get up to 210 minutes. Each relocation estimates the position of the animal, and a minimum of two consecutive relocations are necessary to estimate a trajectory, direction, and velocity. B) Distribution of relative angles for each individual. Relative angles require a minimum of three consecutive relocations to estimate the change in direction.

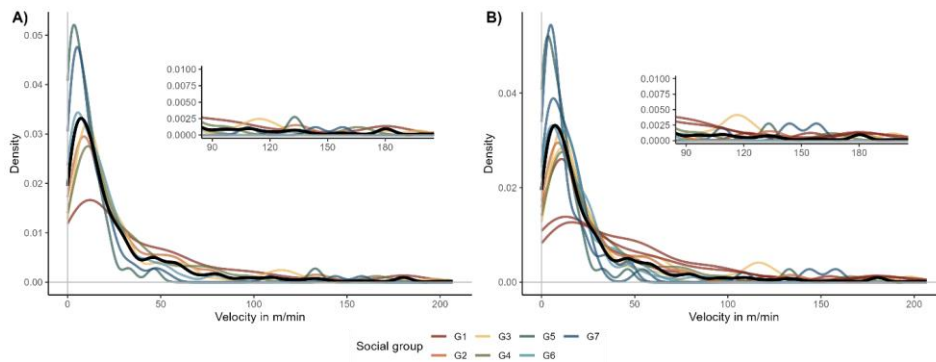


Figure 3-2. Kernel density estimates for the distribution of velocity data at different pooling levels, with the dark black density curve showing complete pooling. A) Partial pooling representation of the data, where each curve displays the density of velocities across a social group. B) No pooling scenario, where each density curve is associated to a single individual's velocity data. Individuals belonging to the same social group share the line color. Plot inserts show a zoom to the tail of the distribution of velocity data in both panels. We observe considerable variation for both partial and no pooling scenarios, when compared to the single black density curve that represents the complete pooling. Particular focus towards the tails, where certain individuals possess much higher densities for large velocities, something that gets dampened when considering pooled data for velocity densities.

Table 3-1. BIC values obtained for models fitted to velocity data across pooling levels. Probability density functions considered included the Gamma, Weibull, and Lognormal distributions. Multi distribution models selected the best fitting distribution model for each social group or individual, allowing for variation not only in estimated parameters, but also in the distribution used. Based on BIC values, the best fitting model is the Multi-distribution with Partial Pooling, followed by the Lognormal PP, and Lognormal CP.

<i>Model</i>	<i>Distribution</i>	<i>BIC</i>	<i>ΔBIC within</i>	<i>ΔBIC across</i>
<i>CP</i>	Gamma	5709.9	25.9	53.3
	Lognormal	5684.0	0	27.5
	Weibull	5698.5	14.5	41.9
<i>PP</i>	Gamma	5699.6	43	43
	Lognormal	5677.1	20.5	20.5
	Weibull	5694.4	37.8	37.8
	Multi	5656.6	0	0
<i>NP</i>	Gamma	5753.5	40.3	96.9
	Lognormal	5732.9	19.8	76.3
	Weibull	5748.6	35.5	92
	Multi	5713.1	0	56.6

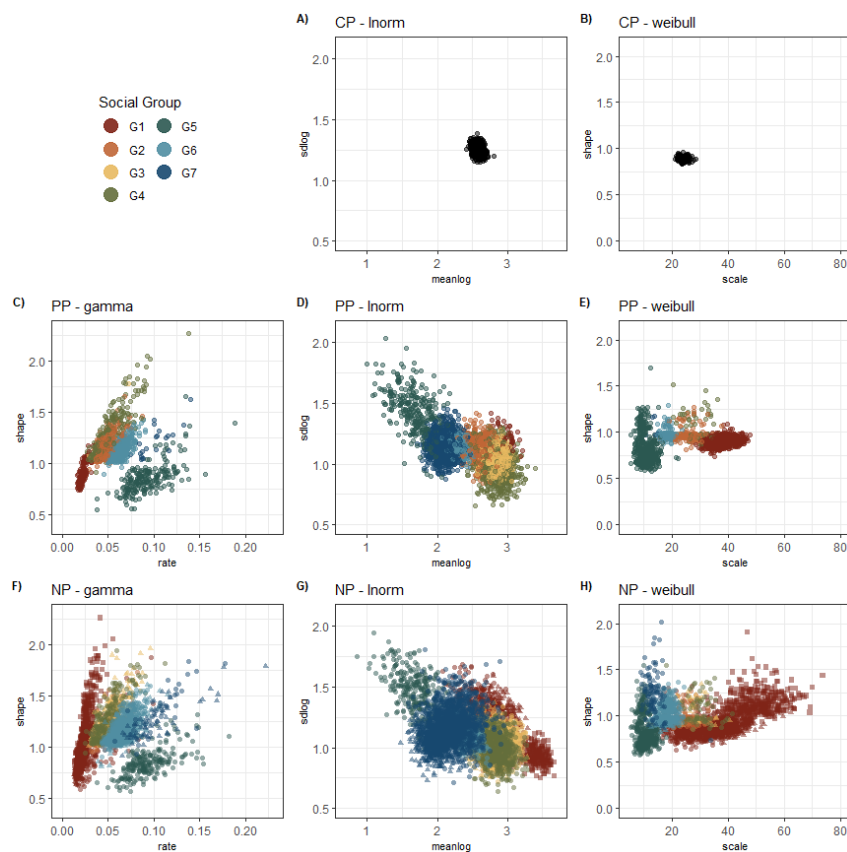


Figure 3-3. Parameters for best fitting models under the three pooling scenarios for 1000 bootstrap replicas. Complete pooling approaches showed best fitting models under the lognormal (panel A.) and weibull (B.) distributions, but no models under a gamma distribution. C)-E) Panels show placement of best fitting models under the three distributions with each color representing a different social group. Under the NP scenario (panels F.-H.) individuals belonging to the same social group are represented with a different shape, but the same color. Every dot in these panels represents only the model with the lowest BIC for that particular bootstrap.

Commented [PCM7]: Each "cloud" of points of different colors represents a confidence "cloud" or region

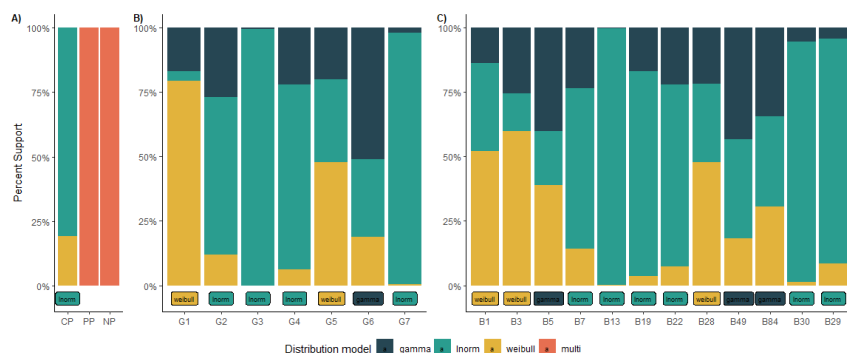


Figure 3-4. Percent support for the different distribution models at the three pooling levels. The percent support is based on 1000 bootstrapped replicas for each scenario. Labels on the x axis show the best fitting model for the original datasets. A) percent support at the three pooling levels, with CP having over 50% support for the lognormal distribution, in agreement with the model fit to the complete dataset. PP and NP scenarios show 100% support for the multi-distribution model, which considers the best fitting model for each family group or individual, respectively. B) Percent support for each submodel in the family groups. C) Support for each submodel at the individual level.

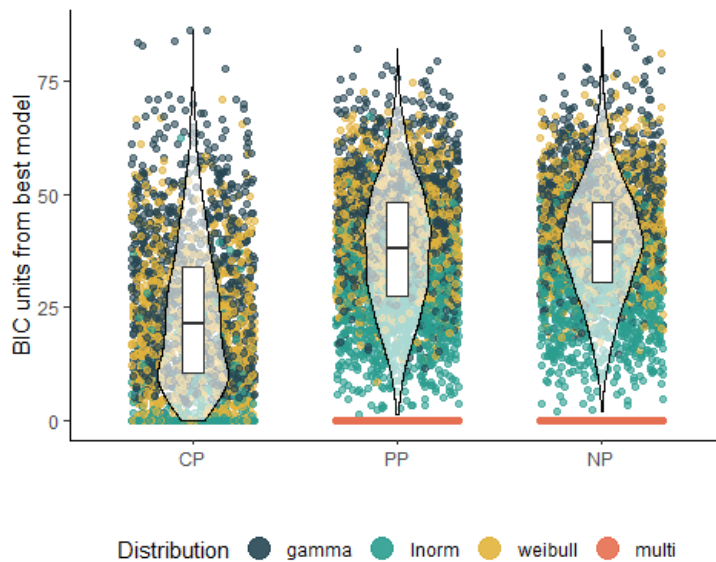


Figure 3-5. Distributions of ΔBIC values for models under the three different pooling scenarios. Best fit model is aligned at $y = 0$, thus in the case of the PP and NP scenario, the orange bar at $y=0$ is actually the ΔBIC value being zero for the multi-distribution model for all bootstraps. The violin plots describe the density of ΔBIC units away from the best model. The box overlay shows the median, and lower and upper quartiles of this distribution for the three different pooling levels.

Commented [PCM8]: The distribution of the evidence favoring the models with heterogeneity (multi) in the cases of PP and NP is much stronger (i.e. much better separated from 0) (thus much stronger) than the distribution of the evidence favoring the lognormal model. (here evidence is the Delta BIC)

Table 3-2. Summary values of the distance between the best model, with BIC = 0, to the second-best model, for all bootstrap replicas across the three pooling scenarios.

Model	minimum	median	mean	q2.5	q97.5
CP	0.05	15.45	18.44	0.73	50.50
PP	1.49	24.10	24.36	6.86	44.10
NP	2.17	27.60	26.93	9.18	43.40

Commented [PCM9]: This result is very important because it says that in the case of PP and NP, the 2.5 quantiles of the evidence function are already showing strong evidence, as opposed to the case of the CP, which shows very weak evidence at the 2.65 quantile.

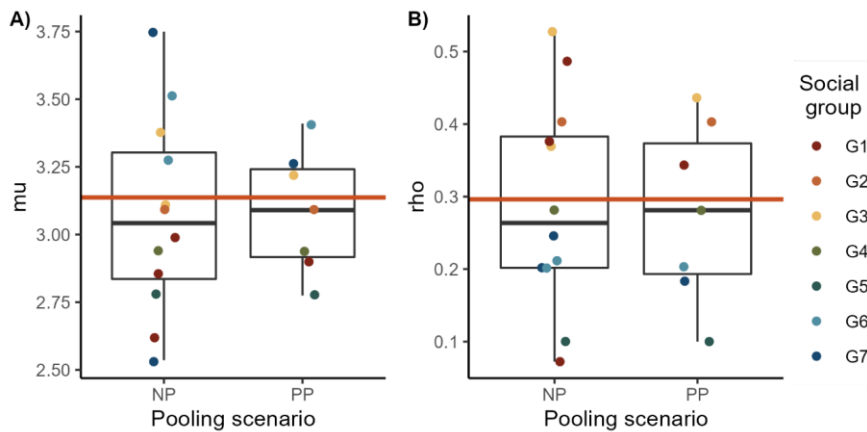


Figure 3-6. Parameters estimated for a wrapped Cauchy distribution for relative angles for individuals, and social groups across the different pooling levels. Color is used to describe the social groups to which individuals belong to. The orange line on both plots represents the parameter values estimated using all the data under the complete pooling approach.

References

- Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, J. Mills Flemming, A. Nielsen, G. Petris, and L. Thomas. 2021. A guide to state–space modeling of ecological time series. *Ecological Monographs* 91:e01470.
- Beckman, N. G., C. E. Aslan, H. S. Rogers, O. Kogan, J. L. Bronstein, J. M. Bullock, F. Hartig, J. HilleRisLambers, Y. Zhou, D. Zurell, J. F. Brodie, E. M. Bruna, R. S. Cantrell, R. R. Decker, E. Efiom, E. C. Fricke, K. Gurski, A. Hastings, J. S. Johnson, B. A. Loiselle, M. N. Miriti, M. G. Neubert, L. Pejchar, J. R. Poulsen, G. Pufal, O. H. Razafindratsima, M. E. Sandor, K. Shea, S. Schreiber, E. W. Schupp, R. S. Snell, C. Strickland, and J. Zambrano. 2020. Advancing an interdisciplinary framework to study seed dispersal ecology. *AoB PLANTS* 12:plz048.
- Bialozyt, R., S. Flinkerbusch, M. Niggemann, and E. W. Heymann. 2014. Predicting the seed shadows of a Neotropical tree species dispersed by primates using an agent-based model with internal decision making for movements. *Ecological Modelling* 278:74–84.
- Borah, B., and N. G. Beckman. 2021. Studying seed dispersal through the lens of movement ecology. *Oikos* n/a.
- Calenge, C. 2006. The package “adehabitat” for the R software: A tool for the analysis of space and habitat use by animals. *Ecological Modelling* 197:516–519.
- Carlo, T. A., and J. M. Morales. 2008. Inequalities in fruit-removal and seed dispersal: consequences of bird behaviour, neighbourhood density and landscape aggregation. *Journal of Ecology* 96:609–618.

- Côrtes, M. C., and M. Uriarte. 2013. Integrating frugivory and animal movement: a review of the evidence and implications for scaling seed dispersal. *Biological Reviews* 88:255–272.
- Dennis, B., J. M. Ponciano, M. L. Taper, and S. R. Lele. 2019. Errors in Statistical Inference Under Model Misspecification: Evidence, Hypothesis Testing, and AIC. *Frontiers in Ecology and Evolution* 7.
- Dray, S., M. Royer-Carenzi, and C. Calenge. 2010. The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research* 25:673–681.
- González-Varo, J. P., and A. Traveset. 2016. The Labile Limits of Forbidden Interactions. *Trends in Ecology & Evolution* 31:700–710.
- Holbrook, K. M. 2011. Home Range and Movement Patterns of Toucans: Implications for Seed Dispersal. *Biotropica* 43:357–364.
- Holbrook, K. M., and B. A. Loiselle. 2007. Using toucan-generated dispersal models to estimate seed dispersal in Amazonian Ecuador. Pages 300–321 *Seed Dispersal: Theory and its Application in a Changing World*. CAB International, Wallingford.
- Holyoak, M., and W. C. Wetzel. 2020. Variance-Explicit Ecology: A Call for Holistic Study of the Consequences of Variability at Multiple Scales. Pages 25–42 *Unsolved Problems in Ecology*. Princeton University Press.
- Hooten, M. B. 2017. *Animal Movement: Statistical Models for Telemetry Data*. CRC Press/Taylor & Francis Group.
- Johnson, D. S., J. M. London, M.-A. Lea, and J. W. Durban. 2008. Continuous-Time Correlated Random Walk Model for Animal Telemetry Data. *Ecology* 89:1208–1215.

- Jones, L. R., S. M. Duke-Sylvester, P. L. Leberg, and D. M. Johnson. 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., J. M. Flemming, and R. A. Myers. 2005. Robust State–Space Modeling of Animal Movement Data. *Ecology* 86:2874–2880.
- Jonsen, I. D., R. A. Myers, and J. M. Flemming. 2003. Meta-Analysis of Animal Movement Using State-Space Models. *Ecology* 84:3055–3063.
- Jonsen, I. D., R. A. Myers, and M. C. James. 2006. Robust hierarchical state–space models reveal diel variation in travel rates of migrating leatherback turtles. *Journal of Animal Ecology* 75:1046–1057.
- Joo, R., S. Picardi, M. E. Boone, T. A. Clay, S. C. Patrick, V. S. Romero-Romero, and M. Basille. 2020. A decade of movement ecology. *arXiv:2006.00110 [q-bio]*.
- Jordano, P., C. García, J. A. Godoy, and J. L. García-Castaño. 2007. Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings of the National Academy of Sciences* 104:3278–3282.
- Kareiva, P. M., and N. Shigesada. 1983. Analyzing insect movement as a correlated random walk. *Oecologia* 56:234–238.
- Karubian, J., and R. Durães. 2009. Effects of seed disperser social behavior on patterns of seed movement and deposition. *Oecologia Brasiliensis* 13:45–57.
- Langrock, R., R. King, J. Matthiopoulos, L. Thomas, D. Fortin, and J. M. Morales. 2012. Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology* 93:2336–2342.

- Lele, S. R. 2020a. How Should We Quantify Uncertainty in Statistical Inference? *Frontiers in Ecology and Evolution* 8.
- Lele, S. R. 2020b. Consequences of Lack of Parameterization Invariance of Non-informative Bayesian Analysis for Wildlife Management: Survival of San Joaquin Kit Fox and Declines in Amphibian Populations. *Frontiers in Ecology and Evolution* 7.
- Lenz, J., W. Fiedler, T. Caprano, W. Friedrichs, B. H. Gaese, M. Wikelski, and K. Böhning-Gaese. 2011. Seed-dispersal distributions by trumpeter hornbills in fragmented landscapes. *Proceedings of the Royal Society B: Biological Sciences* 278:2257–2264.
- Levey, D. J., B. M. Bolker, J. J. Tewksbury, S. Sargent, and N. M. Haddad. 2005. Effects of Landscape Corridors on Seed Dispersal by Birds. *Science* 309:146–148.
- Levey, D. J., J. J. Tewksbury, and B. M. Bolker. 2008. Modelling long-distance seed dispersal in heterogeneous landscapes. *Journal of Ecology* 96:599–608.
- Matthiopoulos, J., J. Fieberg, G. Aarts, H. L. Beyer, J. M. Morales, and D. T. Haydon. 2015. Establishing the link between habitat selection and animal population dynamics. *Ecological Monographs* 85:413–436.
- Morales, J. M., and T. A. Carlo. 2006. The Effects of Plant Distribution and Frugivore Density on the Scale and Shape of Dispersal Kernels. *Ecology* 87:1489–1496.
- Morales, J. M., D. T. Haydon, J. Frair, K. E. Holsinger, and J. M. Fryxell. 2004. Extracting More Out of Relocation Data: Building Movement Models as Mixtures of Random Walks. *Ecology* 85:2436–2445.

- Morales, J. M., P. R. Moorcroft, J. Matthiopoulos, J. L. Frair, J. G. Kie, R. A. Powell, E. H. Merrill, and D. T. Haydon. 2010. Building the bridge between animal movement and population dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:2289–2301.
- Mueller, T., and W. F. Fagan. 2008. Search and navigation in dynamic environments – from individual behaviors to population distributions. *Oikos* 117:654–664.
- Mueller, T., K. A. Olson, G. Dressler, P. Leimgruber, T. K. Fuller, C. Nicolson, A. J. Novaro, M. J. Bolgeri, D. Wattles, S. DeStefano, J. M. Calabrese, and W. F. Fagan. 2011. How landscape dynamics link individual- to population-level movement patterns: a multispecies comparison of ungulate relocation data. *Global Ecology and Biogeography* 20:683–694.
- Nathan, R. 2008. An emerging movement ecology paradigm. *Proceedings of the National Academy of Sciences* 105:19050–19051.
- Nathan, R., and H. C. Muller-Landau. 2000. Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution* 15:278–285.
- Nield, A. P., R. Nathan, N. J. Enright, P. G. Ladd, and G. L. W. Perry. 2020. The spatial complexity of seed movement: Animal-generated seed dispersal patterns in fragmented landscapes revealed by animal movement models. *Journal of Ecology* 108:687–701.
- Patterson, T. A., M. Basson, M. V. Bravington, and J. S. Gunn. 2009. Classifying movement behaviour in relation to environmental conditions using hidden Markov models. *Journal of Animal Ecology* 78:1113–1123.

- Patterson, T. A., A. Parton, R. Langrock, P. G. Blackwell, L. Thomas, and R. King. 2017. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *AStA Advances in Statistical Analysis* 101:399–438.
- Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. State–space models of individual animal movement. *Trends in Ecology & Evolution* 23:87–94.
- Pegman, A. P. M., G. L. W. Perry, and M. N. Clout. 2017. Exploring the interaction of avian frugivory and plant spatial heterogeneity and its effect on seed dispersal kernels using a simulation model. *Ecography* 40:1098–1109.
- Poisot, T., D. B. Stouffer, and D. Gravel. 2015. Beyond species: why ecological interaction networks vary through space and time. *Oikos* 124:243–251.
- R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- Rehm, E., E. Fricke, J. Bender, J. Savidge, and H. Rogers. 2019. Animal movement drives variation in seed dispersal distance in a plant–animal network. *Proceedings of the Royal Society B: Biological Sciences* 286:20182007.
- Rehm, E. M., J. Chojnacki, H. S. Rogers, and J. A. Savidge. 2018. Differences among avian frugivores in seed dispersal to degraded habitats. *Restoration Ecology* 26:760–766.
- Russo, S. E., and C. K. Augspurger. 2004. Aggregated seed dispersal by spider monkeys limits recruitment to clumped patterns in *Virola calophylla*. *Ecology Letters* 7:1058–1067.

- Russo, S. E., S. Portnoy, and C. K. Augspurger. 2006. Incorporating Animal Behavior into Seed Dispersal Models: Implications for Seed Shadows. *Ecology* 87:3160–3174.
- Sasal, Y., and J. M. Morales. 2013. Linking frugivore behavior to plant population dynamics. *Oikos* 122:95–103.
- Schupp, E. W., P. Jordano, and J. M. Gómez. 2017. A general framework for effectiveness concepts in mutualisms. *Ecology Letters* 20:577–590.
- Schupp, E. W., R. Zvolak, L. R. Jones, R. S. Snell, N. G. Beckman, C. Aslan, B. R. Cavazos, E. Effiom, E. C. Fricke, F. Montaña-Centellas, J. Poulsen, O. H. Razafindratsima, M. E. Sandor, and K. Shea. 2019. Intrinsic and extrinsic drivers of intraspecific variation in seed dispersal are diverse and pervasive. *AoB PLANTS* 11.
- Shaw, A. K. 2020. Causes and consequences of individual variation in animal movement. *Movement Ecology* 8:12.
- Skalski, G. T., and J. F. Gilliam. 2000. Modeling Diffusive Spread in a Heterogeneous Population: A Movement Study with Stream Fish. *Ecology* 81:1685–1700.
- Snell, R. S., N. G. Beckman, E. Fricke, B. A. Loiselle, C. S. Carvalho, L. R. Jones, N. I. Lichti, N. Lustenhouwer, S. J. Schreiber, C. Strickland, L. L. Sullivan, B. R. Cavazos, I. Giladi, A. Hastings, K. M. Holbrook, E. Jongejans, O. Kogan, F. Montaña-Centellas, J. Rudolph, H. S. Rogers, R. Zvolak, and E. W. Schupp. 2019. Consequences of intraspecific variation in seed dispersal for plant demography, communities, evolution and global change. *AoB PLANTS* 11.

- Spiegel, O., S. T. Leu, C. M. Bull, and A. Sih. 2017. What's your move? Movement as a link between personality and spatial dynamics in animal populations. *Ecology Letters* 20:3–18.
- Taper, M. L., S. R. Lele, J. M. Ponciano, B. Dennis, and C. L. Jerde. 2021. Assessing the Global and Local Uncertainty of Scientific Evidence in the Presence of Model Misspecification. *Frontiers in Ecology and Evolution* 9:668.
- Turchin, P. 1998. *Quantitative Analysis of Movement: Measuring and Modeling Population Redistribution in Animals and Plants*. Sinauer.
- Violle, C., B. J. Enquist, B. J. McGill, L. Jiang, C. H. Albert, C. Hulshof, V. Jung, and J. Messier. 2012. The return of the variance: intraspecific variability in community ecology. *Trends in Ecology & Evolution* 27:244–252.
- Westcott, D. A., J. Bentrupperbäumer, M. G. Bradford, and A. McKeown. 2005. Incorporating patterns of disperser behaviour into models of seed dispersal and its effects on estimated dispersal curves. *Oecologia* 146:57–67.
- Will, H., and O. Tackenberg. 2008. A mechanistic simulation model of seed dispersal by animals. *Journal of Ecology* 96:1011–1022.
- Zwolak, R. 2018. How intraspecific variation in seed-dispersing animals matters for plants. *Biological Reviews* 93:897–913.
- Zwolak, R., and A. Sih. 2020. Animal personalities and seed dispersal: A conceptual review. *Functional Ecology* 34:1294–1310.

