Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Statistical Modeling of Control of Animal Motion in a Fluid Environment

David Spade

spade@uwm.edu

University of Wisconsin–Milwaukee

J.R. Strickler

University of Wisconsin–Milwaukee

Research Article

Keywords:

Posted Date: August 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9838075/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Statistical Modeling of Control of Animal Motion in a Fluid Environment

David A. Spade[1*]
J.R. Strickler[1]
[1] University of Wisconsin–Milwaukee
Milwaukee, WI, USA

Abstract

This manuscript gets ahold of the control aspect of how *Daphnia magna* move through their surroundings in a control-type environment with still water and with no external stimuli. We present a statistical model for step lengths and roll, pitch, and yaw angles. We then use this model to describe a procedure for generating biologically realistic synthetic trajectories. The aim of this article is to provide a starting point for modeling animal motion from a control-based standpoint in a variety of environmental settings.

1 Introduction

Animals move through their environment for a number of reasons, including to find mates, to forage for food, and to avoid predation. For several decades, biologists have aimed to understand this movement. For animals that move primarily in two-dimensional space, one way to model the movement of animals is to examine how the positional xy coordinates change over time. Early approaches involved the random walk (RW) (Pearson, 1905; Lin and Segel, 1974; Okubo, 1980). This model considers the moves in a seqeunce uncorrelated and unbiased. Komin et al. (2004) applied this model to *Daphnia magna* (Straus, 1820) motion. Patlak (1953) extended this model to allow for correlation between adjacent moves in the random walk through the correlated random walk (CRW). This has since been used by Kareiva and Shigesada (1983), Bovet and Benhamou (1988), and Benhamou (2004). This description of animal motion is more realistic than the RW model, allowing for the notion that in the short term, animals are more likely to continue their direction of movement rather than make large turns. In many cases, an animal will show a consistent preference for a particular direction. This represents a bias in the animal's movement, and this bias should be accounted for. Marsh and Jones (1988) did this with a biased random walk (BRW). This model takes into account the bias in direction, but leaves out the ability to account for correlation among steps. To remedy the lack of accounting for short-term directional preferences, several authors (Benhamou, 2006; Codling et al., 2008; Bailey et al., 2018) employed the biased correlated random walk (BCRW) model, which combines key features of the CRW and BRW models. Later on, several studies were able to use the CRW to model the movement of a variety of species, including beetles (Byers, 2001), elk (Morales et al., 2004; Fortin et al., 2005), and seals (McClintock et al., 2012).
Another common approach to describing movement in two-dimensional space is modeled based on turning angles and step lengths. Broadbent and Kendall (1953) used this approach and

*Corresponding Author: spade@uwm.edu

a random walk model to describe the locomotive behavior of parasitic worm larvae. Kareiva and Shigesada (1983) introduced a two-dimensional CRW to represent the evolution of step lengths and turning angles. Wehner et al. (1992) was able to describe how desert ants, who take a random path to their food source, are able to integrate out the direction and head directly back to the nest after foraging. Komin et al. (2004) used a two-dimensional random walk to ascertain a preferred turning direction of *Daphnia*. Garcia et al. (2007) used turning angle distributions to examine the optimality of *Daphnia* foraging behavior. Gómez-Laich et al. (2018) demonstrated that Magellanic penguins behave in a similar way to how the desert ants behave, taking a somewhat random path from the colony while foraging, but taking a direct path back to it. These penguins forage on the return trip to the colony. For other examples of studies on the movement of animals in a two-dimensional environment, see Craighead et al. (1963); Craighead and Craighead (1965); Deneubourg et al. (1990); Davenport and Bels (2023).
While terrestrial organisms move largely in two-dimensional space , ignoring topography, air and water organisms move in three dimensions. The animal of interest in this work, a planktonic microcrustacean, moves this way. Several techniques are available to model three-dimensional movement. One way is to model positional xyz coordinates directly. Keiyu et al. (1994) used a CRW model to do this. Spade et al. (2024) used a more complex time series model to describe *D. magna* motion over small time scales. Modeling locomotion in this way takes the perspective of the observer. It is possible to minimize this concern by using spherical coordinates (polar angle, azimuth angle, and step length) to model three-dimensional motion (Benhamou, 2019; Ahmed et al., 2021; Benhamou, 2023), but what this misses is the fact that animals can rotate in several different ways. It necessarily treats the agent as a particle without regard for orientation of the animal. Hidden Markov models (HMM) are common choices for modeling animal motion. Whoriskey et al. (2017) constructed an HMM for identifying behavioral states from animal tracks, for example. Many other examples of using HMMs with regard to animal motion exist in the literature. For more, see Ephraim (2002); Cappé et al. (2005); Zucchini et al. (2016); Klappstein et al. (2023); Dupont et al. (2025). Step selection functions have become a common way to represent how animal motion plays its role in habitat selection. These types of models examine the preferences of the agent for particular habitats by looking at the spatial features of locations that have been observed to the features of the surrounding habitat. These have been explored in depth in the literature to date. For more, see Nicosia et al. (2017), Chatterjee et al. (2024), and Klappstein et al. (2024).
With the multitude of methods of describing animal motion in the literature, the common thread is the perspective taken. These methods all use the perspective of the observer to varying degrees. We put the point of reference on the animal by considering the six degrees of freedom of motion. To do this, we think of aviation. An aircraft can move forward (surge), left and right (sway), and up and down (heave). In addition, there are angles to consider. An aircraft can rotate left and right about a vertical axis (yaw), up and down around a lateral axis (pitch), and side-to-side about its own longitudinal axis (roll). Our study animal can do the same things. Essentially, we are placing ourselves in the "cockpit" of the animal and use this point of view to describe the movement of *D. magna*. In so doing, we get at the control aspect of the animal's movement. We do this with no external stimuli in calm water to gain a sense of how the agent moves in a control-type setting. From here, we can add stimuli such as food and predator scent to the water, and we can perturb the water, to examine and describe motion of the agent under different settings, but we do not take this up in this work.
Thinking about animal motion in this way has rarely been explored in the literature. One example of where it has been taken up is in Benhamou (2023). That work proposes using gyroscopic data to estimate roll, pitch, and yaw angles. Their approach uses kernel density estimation to approximate the distribution of these rotation angles. While this works for agents such as birds, our organisms are much too small for measurement of rotation angles in this way.

Therefore, we must use positional coordinates to extract rotation angles and step lengths. The way we do this is described in Section 2.1. In this work, we present a hierarchical model (Pedersen et al., 2019) for rotation angles and step lengths that is used to simulate realistic *D. magna* trajectories. In doing this, we demonstrate how to return the rotation angles and step lengths to positional coordinates for plotting and analytical purposes.
The remainder of this manuscript is organized as follows. Section 2 provides a description and justification of our statistical model of *D. magna* motion, as well as our method for simulating realistic trajectories. We demonstrate the biological realism of the trajectories generated by our simulation procedure in Section 3. The manuscript concludes with a discussion of the results presented herein, along with an outline of future directions of this work.

2 Methods

This section describes our method of collecting a *D. magna* track, the way we analyzed it, and how we used that analysis to inform a useful and versatile probabilistic model of their motion. We then outline the simulation procedure that was born from the model we constructed. This procedure leads to biologically realistic synthetic trajectories.

2.1 Data Collection and Analysis

Seven *D. magna* tracks of different lengths were acquired through the setup described in Spade et al. (2024). These tracks result from recording a position every 1/30 second, but we want to examine positions taken every 1/5 of a second. We thus make use only of every sixth frame. This causes an issue with data availability because the remaining small tracks are not long enough to carry out an effective statistical analysis. Thus, we shift the last six tracks in such a way that the end of one track matches up with the beginning of the next one. We then drop the one overlapping observation to avoid repeated positions where they do not belong. This process creates one long, stitched-together track with a position recorded every 1/5 second and lasting 135.2 seconds, or 676 sequential positions. We recognize that we lose information about what happens in the intervening times between time steps we use, but the actions in those intervening steps are often very small, and would result in heavily zero-inflated distributions of step lengths and rotation angles. This complicates analysis monumentally, while the time step length that we use here alleviates this concern and gives us a track based on which we can build a useful model of zooplankton movement.
This long track is used to inform our model-building process. We first look at the coordinate processes. There is substantial autocorrelation in each of the x, y, and z series, but little partial autocorrelation. Partial autocorrelation represents the correlation of a series with its own lagged values. However, it differs from autocorrelation in the sense that it isolates the correlation between the series and its lagged values while removing the correlation created by intermediate observations. The autocorrelation function (ACF) and partial autocorrelation function (PACF) for each series are in Figures 1 and 2.

These plots suggest that each of the X, Y, and Z series can be modeled with some sort of autoregressive integrated moving average (ARIMA) process. This is indeed the case. Let v_t be a time series process, let $\Delta v_t = v_t - v_{t-1}$, and let $\Delta^2 v_t = \Delta v_t - \Delta v_{t-1} = v_t - 2v_{t-1} + v_{t-2}$. The X series can be modeled with an ARIMA process of orders 1 and 2 with two differences. This is represented by

$$\Delta^2 x_t = \phi \Delta^2 x_{t-1} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2},$$

where the e_t are independent and identically distributed normal error terms with mean 0 and variance ϵ^2. The estimated parameters of the model are $\phi = -0.8439$, θ_1 = -0.5789, and θ_2 =

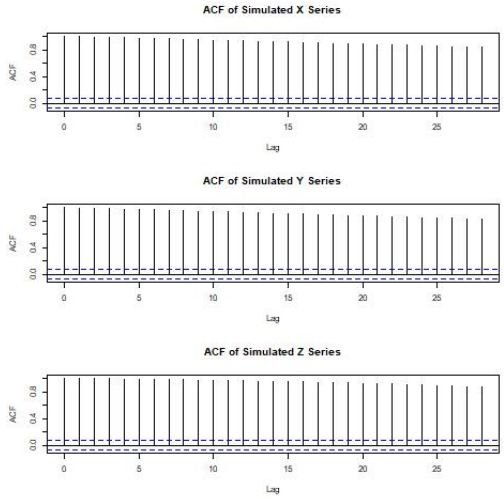


Figure 1: ACF of X (top), Y (middle), and Z (bottom) series.

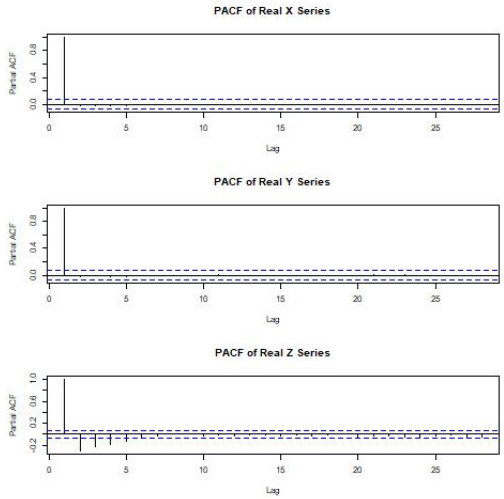


Figure 2: PACF of X (top), Y (middle), and Z (bottom) series.

0.2750. The Y series can be represented with an ARIMA process of orders 4 and 1 with two differences. This yields

$$\Delta^2 y_t = \phi_1 \Delta^2 y_{t-1} + \phi_2 \Delta^2 y_{t-2} + \phi_3 \Delta^2 y_{t-3} + \phi_4 \Delta^2 y_{t-4} + e_t + \theta e_{t-1}.$$

The estimated parameters of this process are $\phi_1 = 0.5351$, $\phi_2 = 0.3003$, $\phi_3 =$ -0.0303, $\phi_4 =$ -0.1575, and $\theta = 0.9687$. The Z series can be represented by an ARIMA process of orders 3 and 2 with one difference, represented by

$$\Delta z_t = \phi_1 \Delta z_{t-1} + \phi_2 \Delta z_{t-2} + \phi_3 \Delta z_{t-3} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2},$$

where the same assumptions apply for the e_t series in the Y and Z processes as applied for the X series. The estimates of the parameters for this model are $\phi_1 = 1.2152$, $\phi_2 =$ -0.4709, $\phi_3 =$ 0.1352, $\theta_1 = 0.9648$, and $\theta_2 =$ -0.4492.

These processes, while useful in describing the positional coordinates of the animal under observation, are severely limited in several ways. The primary limitation is that, due to the specificity of the model, it would be difficult to say confidently that we can apply this to other animals, even within the same species, simply by changing the order of the model. It would likely be necessary to change the orders of the model, and perhaps even the number of

differences taken for each series. The fact that these processes describe positional coordinates means that these are reactive models. We observe positional coordinates, and then describe them. This takes the viewpoint of the observer. What we aim to do is to take the viewpoint of the agent by putting the point of reference on it. We make use of the control aspect through constructing models for step lengths and rotation angles. This is all missed by using the models we just described.

At this point, we describe our analysis of the observed track and demonstrate the correctness of our work by showing that reversing the mathematical process recovers the "true" track. Obtaing the step lengths r_t, $t = 1, \ldots, 673$ is a straightforward process. We have

$$r_t = \sqrt{((\Delta x_t)^2 + (\Delta y_t)^2 + (\Delta z_t)^2)}.$$

To obtain roll, pitch, and yaw angles, we use a mathematical object called a *quaternion*. A quaternion is a four-dimensional representation of three-dimensional space. Let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ represent the unit vectors in the x, y, and z directions, respectively. Let a, b, c, and d denote scalar constants. A quaternion $\langle a, b, c, d\rangle$ represents the vector $a\mathbf{i} + b\mathbf{j} + c\mathbf{k} + d$ in our setup. To construct the quaternions for use in our approach, we let $\mathbf{u}_1$ and $\mathbf{u}_2$ be two three-dimensional unit vectors. Define the *dot product* of $\mathbf{u}_1$ and $\mathbf{u}_2$ as

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = \sum_{i=1}^{3} u_{1i} u_{2i}.$$

Next, define the angle θ between $\mathbf{u}_1$ and $\mathbf{u}_2$ to be

$$\theta = \cos^{-1}(\mathbf{u}_1 \cdot \mathbf{u}_2).$$

We now define the cross product $\mathbf{w}$ between $\mathbf{u}_1$ and $\mathbf{u}_2$. This product gives a vector that is perpendicular to both $\mathbf{u}_1$ and $\mathbf{u}_2$. Define

$$\begin{aligned} \mathbf{w} = \mathbf{u}_1 \times \mathbf{u}_2 &= (\mathbf{u}_{12}\mathbf{u}_{23} - \mathbf{u}_{13}\mathbf{u}_{22})\mathbf{i} \\ &- (\mathbf{u}_{11}\mathbf{u}_{23} - \mathbf{u}_{13}\mathbf{u}_{21})\mathbf{j} \\ &+ (\mathbf{u}_{11}\mathbf{u}_{22} - \mathbf{u}_{12}\mathbf{u}_{21})\mathbf{k}. \end{aligned}$$

Our quaternion is defined as a function of $\mathbf{u}_1$ and $\mathbf{u}_2$, as well as θ, implicitly. If we let $\mathbf{w}$ denote the cross product between $\mathbf{u}_1$ and $\mathbf{u}_2$, we have the following: if the length $\|\mathbf{w}\|$ of $\mathbf{w}$ is 0, then

$$\mathbf{q}(\mathbf{u}_1, \mathbf{u}_2) = \langle 0, 0, 0, 1\rangle .$$

Otherwise,

$$\mathbf{q}(\mathbf{u}_1, \mathbf{u}_2) = \left\langle \mathbf{w} \sin\left(\frac{\theta}{2}\right), \cos\left(\frac{\theta}{2}\right) \right\rangle .$$

To get the roll, pitch, and yaw angles for a given time step, we find the quaternion $\mathbf{q}$ for the direction vectors at the previous and current time step. We then obtain the following rotation matrix:

$$\mathbf{R}(\mathbf{q}) = \begin{bmatrix} 1 - 2(\mathbf{q}_2^2 + \mathbf{q}_3^2) & 2(\mathbf{q}_1\mathbf{q}_2 - \mathbf{q}_3\mathbf{q}_4) & 2(\mathbf{q}_1\mathbf{q}_3 + \mathbf{q}_2\mathbf{q}_4) \\ 2(\mathbf{q}_1\mathbf{q}_2 + \mathbf{q}_3\mathbf{q}_4) & 1 - 2(\mathbf{q}_1^2 + \mathbf{q}_3^2) & 2(\mathbf{q}_2\mathbf{q}_3 - \mathbf{q}_1\mathbf{q}_4) \\ 2(\mathbf{q}_1\mathbf{q}_3 - \mathbf{q}_2\mathbf{q}_4) & 2(\mathbf{q}_2\mathbf{q}_3 + \mathbf{q}_1\mathbf{q}_4) & 1 - 2(\mathbf{q}_1^2 + \mathbf{q}_2^2) \end{bmatrix}.$$

This gives the rotation matrix for the next move. To get the roll (α), pitch (β), and yaw (γ) angles, we use the following, beginning with pitch angle.

$$\beta = \operatorname{atan2}\left(-\mathbf{R}(\mathbf{q})_{3,1}, \sqrt{\mathbf{R}(\mathbf{q})_{1,1}^2 + \mathbf{R}(\mathbf{q})_{2,1}^2}\right),$$

where atan2$(\cdot,\cdot)$ is the two-argument arctangent that puts the angle in the correct quadrant. Once we have a pitch angle, if $\cos(\beta) = 0$ and $\mathbf{R}(\mathbf{q})_{3,1}$ = -1, $\beta = 90°$ and

$$\alpha = \text{atan2}\left(\mathbf{R}(\mathbf{q})_{1,2}, \mathbf{R}(\mathbf{q})_{1,3}\right).$$

In this situation, the yaw angle γ is 0°. If $\cos(\beta) = 0$ and $\mathbf{R}(\mathbf{q})_{3,1} = 1$, $\gamma = 0$, β = -90°, and

$$\alpha = \text{atan2}\left(-\mathbf{R}(\mathbf{q})_{2,1}, -\mathbf{R}(\mathbf{q})_{1,3}\right).$$

In any other situation, we have

$$\begin{aligned} \alpha &= \text{atan2}\left(\frac{\mathbf{R}(\mathbf{q})_{3,2}}{\cos(\beta)}, \frac{\mathbf{R}(\mathbf{q})_{3,3}}{\cos(\beta)}\right) \text{ and} \\ \gamma &= \text{atan2}\left(-\frac{\mathbf{R}(\mathbf{q})_{2,1}}{\cos(\beta)}, \frac{\mathbf{R}(\mathbf{q})_{1,1}}{\cos(\beta)}\right). \end{aligned}$$

If our work is correct, then we should be able to recover the true track through a series of rotation matrices applied to the rotation angles we found in the procedure described above. To begin, let the initial direction vector $\mathbf{d}_1 = \langle 1, 0, 0\rangle$. Let $\mathbf{R}_r(\alpha)$ be the rotation matrix for roll, $\mathbf{R}_r(\beta)$ be the rotation matrix for pitch, and $\mathbf{R}_r(\gamma)$ be the rotation matrix for yaw. These matrices are given by

$$\begin{aligned} \mathbf{R}_r(\alpha) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{bmatrix}, \\ \mathbf{R}_p(\beta) &= \begin{bmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{bmatrix}, \text{ and} \\ \mathbf{R}_y(\gamma) &= \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Then we use the matrix $\mathbf{R}^*(\alpha, \beta, \gamma) = \mathbf{R}_y(\gamma)\mathbf{R}_p(\beta)\mathbf{R}_r(\alpha)$, where the multiplication MUST be taken in this order. To recover a track with the same shape as the observed one, we randomly choose a starting point $\mathbf{x}$. Recall that $\mathbf{d}_1$ is the initial direction. For each subsequent time step, we take $\mathbf{d}_i = \mathbf{R}^*(\alpha, \beta, \gamma)\mathbf{d}_{i-1}$. Then $\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{d}_i * r_{i-1}$. The illustration of the recovery of the track is seen in Figure 3. The blue curves represent the observed x, y, and z series, and the black curves represent the recovered series.

2.2 Model Construction

To construct our model, we handle step lengths and turning angles independently. We recognize that, due to morphology, there may be limitations to this approach. For example, an agent may be able to move farther in a time step when they do not turn as sharply than they can when they make a large turn. However, we will show that this simplifying assumption does not cost us much in terms of biological realism of our simulated trajectories.

2.2.1 Modeling Step Lengths

To represent step lengths, we notice that sometimes the agent makes much larger moves than at other times. We call a large move any move in a single time step that is larger than 25 pixels, where a pixel is 0.113 mm × 0.113 mm. This threshold for large steps is chosen because steps

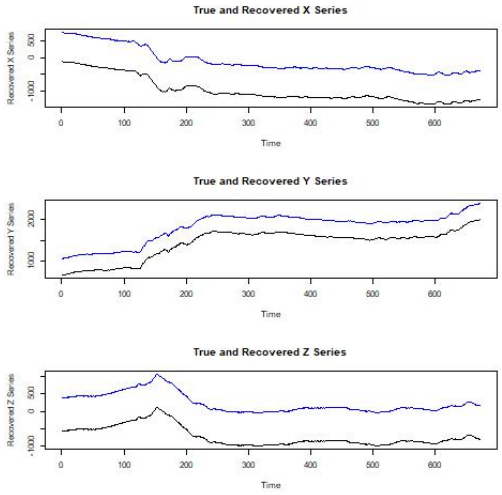


Figure 3: The observed and recovered X (top), Y (middle), and Z (bottom) series. The observed series is drawn in blue, and the recovered track is drawn in black. We see that the recovered and observed series run parallel to each other.

25 pixels or larger only occur about 5% of the time. These larger moves tend to be clustered together in the time series. Let p_{lg} be the proportion of moves that are large. We call moves less than 25 pixels "regular" moves. Assume p_{lg} is distributed according to a Beta distribution with shape parameters α_{lg} and β_{lg}, so that the mean proportion of large moves is $\frac{\alpha_{lg}}{\alpha_{lg}+\beta_{lg}}$. Let n_{lg} denote the number of large moves in a series of length n. We assume n_{lg} follows a binomial distribution with n trials and success probability p_{lg}. The number n_r of regular moves is given by $n_r = n - n_{lg}$, and we observe no moves of length 0. We thus do not model these. While such "moves" are possible, they are rare enough that even in simulating longer series, not accounting for it will not cost us in terms of biological realism.

At this point, we consider the subsequences where the vast majority of large moves occur. In the observed track, we note one such subsequence of the step length series. However, we recognize that there may be more than one such stretch, so we account for this possibility. Let s denote the number of subsequences with a high percentage of large step lengths. Let $\kappa = 1, 2, \ldots, s$ be the indices of these subsequences. We assume that s follows a Poisson distribution with mean λ. Let n_κ denote the number of large steps in each of the s subsequences, and let p_κ denote the proportion of each of these subsequences composed of large steps. Then we assume that the vector $\mathbf{n} = (n_1, \ldots, n_s)$ follows a multinomial distribution with n_{lg} trials and category probability vector $\frac{1}{s}\mathbf{1}_s$, where $\mathbf{1}_s$ is a length s vector of ones. We assume that for $\kappa = 1, 2, \ldots, s$, p_κ follows a Beta distribution with shape parameters α_κ and β_κ. Letting l_κ denote the length of subsequence κ, we have $l_\kappa = \left\lceil \frac{n_\kappa}{p_\kappa} \right\rceil$, where $\lceil x \rceil$ is the smallest integer larger than or equal to x. Letting $\mathbf{b} = \langle b_1, \ldots, b_s \rangle$ be the vector of starting points of each subsequence and $\mathbf{e} = \langle e_1, \ldots, e_l \rangle$ be the vector of ending points of each subsequence, with $e_0 = 0$, b_κ is assumed uniform on the set $\left\{ e_{\kappa-1} + 1, e_{\kappa-1} + 2, \ldots, n - \sum_{j=\kappa}^{s} l_j \right\}$, for i in $1, \ldots, s$. We let $\mathbf{v}_\kappa$ be the vector of time points in subsequence κ where a large move occurs. Given l_κ, we assume $\mathbf{v}_\kappa$ is uniform among the $\binom{l_\kappa}{n_\kappa}$ possible samples of n_κ spots from l_κ. Letting $\mathbf{m}$ be the vector of step lengths, and $\mathbf{v}$ be the vector of all time steps where a large move occurs, we assume $\mathbf{m}_{\mathbf{v}_1}, \ldots, \mathbf{m}_{\mathbf{v}_{n_{lg}}}$ are independent and identically distributed according to a uniform distribution on the interval from 25 to 45 pixels. We choose 45 as an upper bound because we did not observe any step lengths larger than 45 pixels in the *D. magna* track. If $\boldsymbol{\omega}$ is the vector of time points containing regular step lengths, we assume $\mathbf{m}_{\boldsymbol{\omega}_1}, \mathbf{m}_{\boldsymbol{\omega}_2}, \ldots, \mathbf{m}_{\boldsymbol{\omega}_{n_{reg}}}$ be independent and identically distributed from a

chi-square distribution with ν degrees of freedom, where ν is drawn from a Poisson distribution with rate parameter λ_ν. This setup fully describes the model for the step lengths.

2.2.2 Modeling Rotation Angles

To model rotation angles, we first note that there are situations in which no rotation happens, or when only roll, pitch, or yaw happens. We model these probabilities p_N, p_Y, p_P, and p_R, respectively, with beta distributions with parameters estimated from the data as follows:

$$\begin{aligned} p_N &\sim \text{Beta}(6, 994) \\ p_Y &\sim \text{Beta}(1, 99) \\ p_P &\sim \text{Beta}(3, 97) \\ p_R &\sim \text{Beta}(4, 96) \end{aligned}$$

These parameters were selected because no rotation time steps occurred 0.6% of the time in the observed track, and these parameter choices give that value as the mean value of p_{NT}. The parameters of the beta distributions for the other rotation probabilities are selected for similar reasons. To determine the number n_N, n_Y, n_P, and n_R of each type of rotation listed above, we take

$$\begin{aligned} n_N &= \left\lfloor \frac{n}{p_N} \right\rfloor \\ n_Y &= \left\lfloor \frac{n}{p_Y} \right\rfloor \\ n_P &= \left\lfloor \frac{n}{p_P} \right\rfloor \\ n_R &= \left\lfloor \frac{n}{p_R} \right\rfloor, \end{aligned} \tag{1}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. The floor function is chosen here to counteract the ceiling function chosen earlier, thus avoiding the unlikely scenario that we run out of time steps from which to choose. Let $\mathbf{a}_N$ be the vector of time steps in which no turn occurs. We assume that this vector is uniformly distributed among the $\binom{n}{n_N}$ possible samples of n_N spots from n. Similarly, $\mathbf{a}_Y$ is assumed uniformly distributed among the $\binom{n-n_N}{N_Y}$ samples of n_Y time steps from the remaining $n - n_Y$. The vector $\mathbf{a}_P$ of time steps where only a pitch rotation occurs is assumed uniform among the $\binom{n-n_N-n_Y}{n_P}$ possible samples of n_P time steps from the remaining available time slots. Finally, we assume $\mathbf{a}_R$ is uniform among the remaining $\binom{n-n_N-n_Y-n_P}{n_R}$ possible samples of n_R time steps. We now model the rotation angles of each type, and then we model the rotations where all three components are involved. We term these rotations "regular" rotations. Letting $\boldsymbol{\delta}_Y$, $\boldsymbol{\delta}_P$, and $\boldsymbol{\delta}_R$ denote the angles where only a yaw occurs, only a pitch change occurs, and only a roll change occurs, respectively, and letting $TN(a, b, c, d)$ denote the truncated normal distribution with mean a and variance b on the interval from c to d, we assume the following:

$$\begin{aligned} \boldsymbol{\delta}_Y &\overset{iid}{\sim} TN(0, 1.75^2, -\pi, \pi) \\ \boldsymbol{\delta}_P &\overset{iid}{\sim} TN(0, 1.25^2, -\pi/2, \pi/2) \\ \boldsymbol{\delta}_R &\overset{iid}{\sim} TN(0, 1.75^2, -\pi, \pi), \end{aligned}$$

where the expression $\overset{iid}{\sim}$ is taken to mean the entries of the given vector are independent and identically distributed according to the distribution given on the right-hand side. The selected

mean and variance parameters are informed by the data. While mathematically, it causes no challenge to model the special cases and the regular rotations together, we handle them separately because the special cases come up more frequently in the data than what would be borne out by modeling all of them together.
Now let $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ denote the regular roll, pitch, and yaw angles, respectively, we assume the following:

$$\begin{aligned}\boldsymbol{\alpha} &\overset{iid}{\sim} TN(0, 1.25^2, -\pi, \pi)\\ \boldsymbol{\beta} &\overset{iid}{\sim} TN(0, 0.75^2, -\pi/2, \pi/2)\\ \boldsymbol{\gamma} &\overset{iid}{\sim} TN(0, 1.25^2, -\pi, \pi).\end{aligned}$$

The step length and rotation angle distributions were chosen based on the histograms given in Figure 4. Another reason we avoid ARIMA models for the step lengths and the rotation

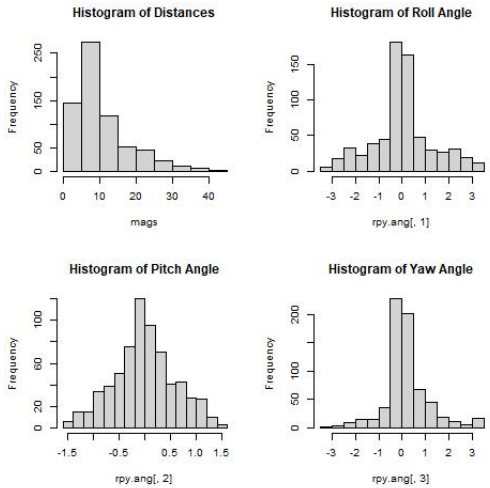


Figure 4: Histograms of step length (Top Left), roll angle (Top Right), pitch angle (Bottom Left), and yaw angle (Bottom Right).

angles is that the ACF and PACF do not really bear out that the gain in capturing variation from an ARIMA model is worth what we lose in terms of generalizability. Thre is considerable autocorrelatoin in the step lengths, but little in the rotation angles. The ACF is shown in Figure 5. We see little significant partial autocorrelation in the rotation angles, and what is significant is small enough to ignore. There is some significant partial autocorrelation in the step lengths, but we do not use this information to select an appropriate ARIMA model. The PACF plots of the observed step lengths and rotation angles are in Figure 6.

2.3 Simulation Procedure

Our method of generating synthetic *D. magna* trajectories is based on the model described in Section 2.2. We generate a series the same length n as the observed one for comparative purposes. We begin with the simulation of step lengths. The proportion p_{lg} of moves that are large is simulated from a Beta distribution with parameters 6 and 94, so that we end up with roughly six percent of step lengths being greater than 25 pixels. The number n_{lg} of large steps is drawn from a binomial distribution with n trials and success probability p_{lg}. Next, we generate the number s of subsequences with large steps from a Poisson distribution with rate parameter $n/700$. If we draw a 0, we add 1 so that we have at least one such subsequence. We then distribute the n_{lg} steps uniformly at random among the s subsequences. For each subsequence, we draw the proportion p_κ of the subsequence κ consisting of large steps from

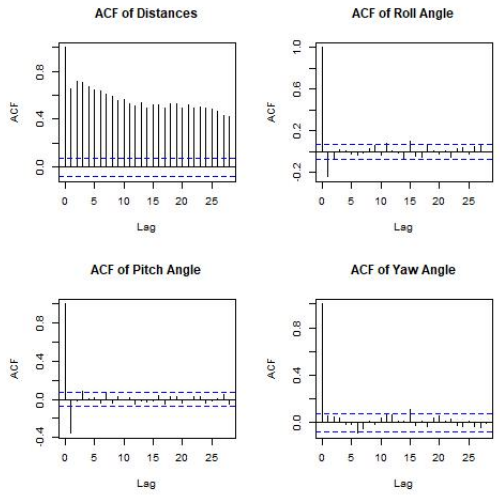


Figure 5: Autocorrelation plots of observed step length (top left), roll angle (top right), pitch angle (bottom left), and yaw angle (bottom right).

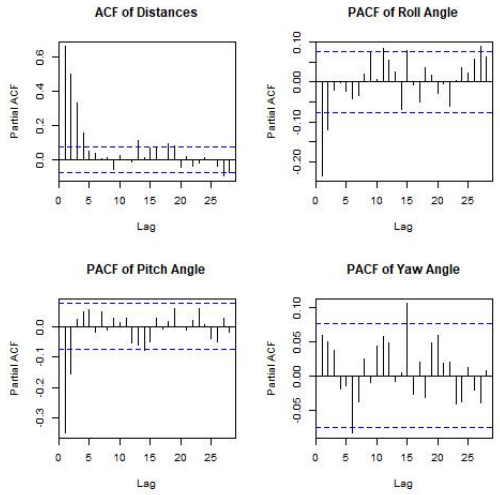


Figure 6: Partial autocorrelation plots of observed step length (top left), roll angle (top right), pitch angle (bottom left), and yaw angle (bottom right).

a Beta distribution with shape parameters 40 and 60. This gives a mean proportion of 0.4. Next, the total length l_κ of subsequence κ is computed as $l_\kappa = \left\lceil \frac{n_\kappa}{p_\kappa} \right\rceil$. The initial times of the s subsequences are drawn uniformly as described in Section 2.2. The large step lengths are drawn independently from a uniform distribution on the interval (25,45), and the regular step lengths are drawn independently from a chi-square distribution with ν degrees of freedom, where ν is drawn from a Poisson distribution with rate parameter 9. This choice was supported by the data, as the average regular step length was near 9 pixels.

To simulate the rotation angles, we begin by choosing our no-turn probability from a Beta distribution with shape parameters 6 and 994. This results in an average probability of 0.006 of no turn at a given time step. This is supported by the data, where approximately 0.6% of the moves were based on no rotation. The data also suggest that 1% of steps happen with only yaw, 3% with only pitch, and 4% happen with only roll. Thus, the yaw only probability is drawn from a Beta distribution with shape parameters 1 and 99, the pitch-only probability is drawn from a Beta distribtution with shape parameters 3 and 97, and the roll-only probability is drawn from a Beta distribution with shape parameters 4 and 96. These result in the average probabilities

matching those borne out by the data. Next, we compute the numbers of rotation-free, yaw-only, pitch-only, and roll-only rotations using the formulas in equations (2). The positions of each type of rotation described above are drawn uniformly at random from the time steps that are available. The yaw-only angles are drawn from a truncated normal distribution on the interval from -π to π radians, with mean 0 and variance 1.75. The roll-only rotation angles are drawn from this distribution as well. The pitch-only rotations are drawn from a truncated normal distribution on the interval from $-\pi/2$ to $\pi/2$ with mean 0 and variance 1.25. Next, the remaining roll angles are drawn independently from a truncated normal distribution on the interval from $-\pi$ to π with mean 0 and variance 1.25. The yaw angles are drawn independently from the same distribution, while the pitch angles are drawn independently from a truncated normal distribution on the interval from $-\pi/2$ to $\pi/2$ with mean 0 and variance 0.75. These choices are all supported by the estimates of the means and variances that come from the data. For plotting purposes, we convert the step lengths and rotation angles back to rectangular coordinates in the manner described in Section 2.2. The resulting tracks end up being less smooth than the observed ones, so we smooth these by taking a rolling mean of order 3. In other words, the smoothed vector $\mathbf{x}_t^*$ is given by

$$\mathbf{x}_t^* = \frac{1}{3}\left(\mathbf{x}_t + \mathbf{x}_{t+1} + \mathbf{x}_{t+2}\right)$$

for $t = 1, 2, \ldots, n-2$. We lose two observations by doing this, but the series are smoother than they were before. We will discuss our measures of smoothness in Section 2.4.

2.4 Quantifying Smoothness

We measure smoothness by the variogram-based estimate of the fractal dimension. Larger values of the fractal dimension indicate a rougher trajectory. To obtain this quantity for the x, y, and z directions, we consider a lag of h time steps. Let $N(h)$ denote the set of all pairs of values in the series separated by h time steps. Let $|N(h)|$ denote the number of elements in this set. For given h, compute the variogram $\gamma(h)$, where

$$\gamma(h) = \frac{1}{2|N(h)|} \sum_{(i,j)\in N(h)} (x_i - x_j)^2.$$

Compute $\gamma(h)$ for $h = 1, 2, \ldots, l$, $l < n$. Regress the natural logarithm of $\gamma(h)$ on $\ln(h)$, and let α denote the slope of the regression line. The variogram estimate D is given by

$$D = \frac{5 - \alpha}{2}.$$

The fractal dimensions of the observed tracks are 1.083 for the x direction, 1.096 for the y direction, and 1.194 for the z direction. We also make use of the mean squared difference of normalized first difference (MSD). To find this for series x_t, we do the following. First, compute the first difference series $\Delta x_t = x_t - x_{t-1}$. Then normalize them to $N\Delta x$ by taking

$$N\Delta x_t = \frac{\Delta x_t - \bar{\Delta x_t}}{\mathrm{SD}(\Delta x_t)},$$

where $\bar{\Delta x_t}$ denotes the mean of the Δx_t series. The roughness r_t is then computed by taking $(N\Delta x_t - N\Delta x_{t-1})^2/4$. Then take the mean of the $N\Delta x_t$. For the real X series, this measure is 0.116. For the real Y series, we have a value of 0.138, and for the real Z series, this measure is 0.238. We look at the lag 1 autocorrelation now as a smoothness measure. For the real X series, this measure is 0.995, for the real Y series, we have a lag one autocorrelation of 0.994, and for the real Z series, we have a value of 0.999.

3 Results

Here, we use several graphical and numerical measures to compare the behavior between the real and simulated trajectories. We note here that we do not expect the simulated tracks to look like the real tracks. We do expect them to behave similarly in terms of step length and rotation angle distributions, autocorrelation and partial autocorrelation, and smoothness. To begin, we examine the time series plots of the resulting positional time series. These are shown in Figure 7. The time series plots of the simulated series show roughness around a slight trend for all

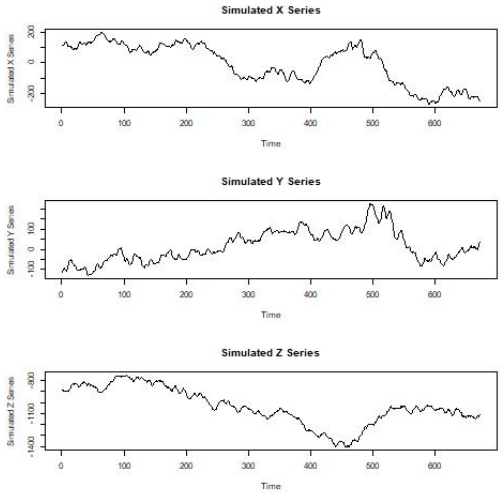


Figure 7: Simulated X series (top), simulated Y series (middle), and simulated Z series (bottom).

three series. This is similar to what we see in the plots of the observed series, though the trends appear to go in other directions. Next, we examine autocorrelation and partial autocorrelation in the simulated x, y, and z series. The autocorrelation plots are in Figure 8. We see that the autocorrelation decreases very slowly in each of the X, Y, and Z coordinates, much like in the observed series. The partial autocorrelation in the X, Y, and Z series is small for lags 2 and

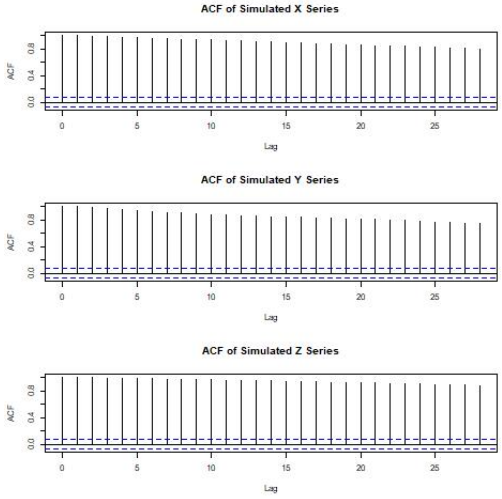


Figure 8: ACF of simulated X series (top), Y series (middle), and Z series (bottom).

on, but significant for lag 1. This is roughly the same behavior as the partial autocorrelation in the observed X, Y, and Z series. The PACFs for the simulated series are shown in Figure 9. We now look at the ACF and PACF plots for the step lengths and the rotation angles. In the

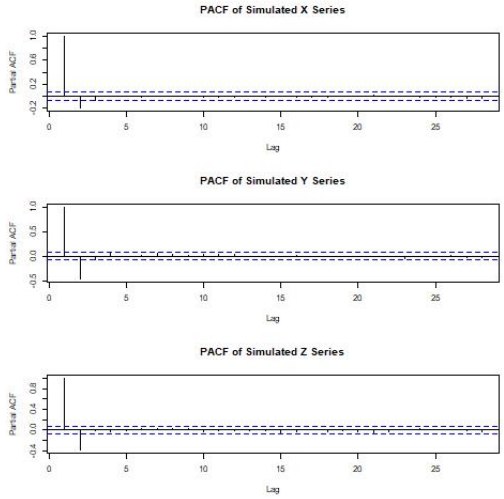


Figure 9: PACF of simulated X series (top), Y series (middle), and Z series (bottom).

ACF plot for the step lengths, we see small but significant autocorrelation for most lags up to 25. This is a deviation from what we see in the observed track, where the step lengths are more strongly autocorrelated. While we missed some of the step length autocorrelation in the simulation procedure, we find that it does not affect the biological realism of our simulated tracks in any meaningful way. This autocorrelation was missed because we modeled step lengths as independent observations. Had we used some sort of an ARIMA model for them, we would have captured this autocorrelation. However, we then run into similar generalizability issues to what we encountered by using this type of model for xyz coordinates.
For the rotation angles, there is little autocorrelation to mention in any of the series pertaining to the simulated roll, pitch, and yaw angles. This is on par with what we see in the observed roll angle, pitch angle, and yaw angle series. The simulation procedure captures the lack of autocorrelation in these series. The ACF plots can be seen in Figure 10. Next, we compare

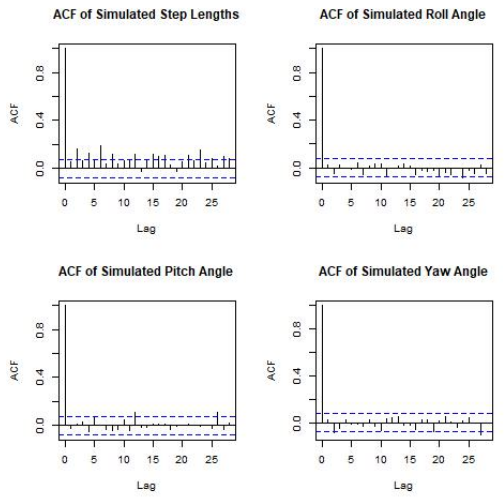


Figure 10: Autocorrelation plots of the simulated step lengths (top left), simulated roll angles (top right), simulated pitch angles (bottom left), and simulated yaw angles (bottom right).

the partial autocorrelation for the simulated step lengths and rotation angles to the partial autocorrelation of the observed step lengths and rotation angles. The simulated step length series shows small but significant partial autocorrelation at several lags. The partial autocorrelation in the simulated roll angle series is small and largely not significant. This is also the

case for the simulated pitch angle and yaw angle series. The partial autocorrelation plots are shown in Figure 11. We see a minor deviation in behavior in this way from the observed series, where there is significant partial autocorrelation in the step length series for small lags. The simulation procedure does not capture this partial autocorrelation well, but this also does not cost us anything in terms of biological realism. The next thing to examine are histograms

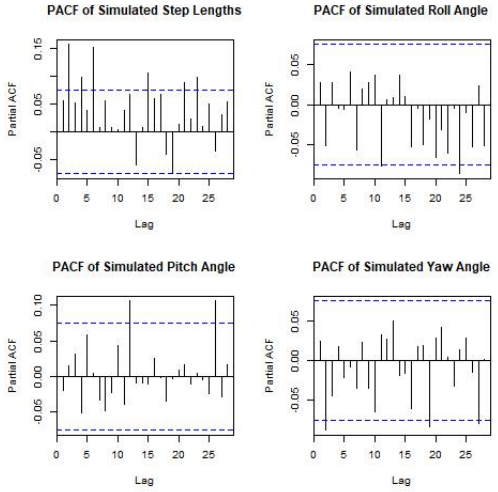


Figure 11: Partial autocorrelation plots of the simulated step lengths (top left), simulated roll angles (top right), simulated pitch angles (bottom left), and simulated yaw angles (bottom right).

of the simulated step lengths and rotation angles. We compare them to the histograms from the observed series and notice similar shapes to all four histograms. The histograms of the simulated rotation angles show less kurtosis (peakedness) than those from the observed track, except for the one corresponding to the pitch angle. The histogram of the pitch angle looks very similar to the histogram based on the observed track. The histograms for the simulated track are in Figure 12. We now address smoothness of our simulated track. We observe that

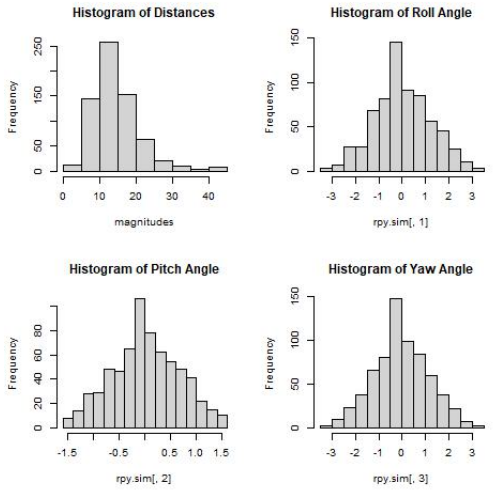


Figure 12: Histograms of simulated step lengths (top left), roll angle (top right), pitch angle (bottom left), and yaw angle (bottom right).

our simulated track is rougher than our observed track. The MSD measures are 0.383 for the simulated X series, 0.474 for the simulated Y series, and 0.402 for the simulated Z series. These

are much higher than those for the observed series. The lag-1 autocorrelations look similar to those for the observed series, with a lag 1 autocorrelation of 0.995 for the X series, 0.992 for the Y series, and 0.995 for the Z series. The estimated fractal dimensions are considerably higher for the simulated series than they are for the observed series. For the simulated X series, the estimated fractal dimension is 1.348, while it is 1.461 and 1.367 for the simulated Y and Z series, respectively. To smooth these series, we take a three-step rolling average to the positional coordinate vector. Doing this gives a new set of simulated X, Y, and Z series. The MSD values for these resulting series are 0.090 for the X series, 0.151 for the Y series, and 0.183 for the Z series. These are much closer to the MSD values obtained from the observed track. The lag-1 autocorrelations continue to be very close to those obtained from the observed track, with 0.996 for the X series, 0.995 for the Y seires, and 0.996 for the Z series. The fractal dimensions come much closer to those from the observed track as well, with values of 1.067 for the X series, 1.116 for the Y series, and 1.143 for the Z series. We also capture the fact that the Z track is a bit rougher than the X and Y series due to the hop-sink behavior of *D. magna*.

4 Conclusion

The goal of this manuscript was to get ahold of the control aspect of how *D. magna* move through their surroundings in a baseline environment. The strategy was to examine the step lengths and the roll, pitch, and yaw angles the agent makes as they move through their environment. We show that the mathematical setup we use to do this is correct by demonstrating that the true track can be recovered by using them in reverse. This takes the perspective completely of the agent, rather than taking the reactive perspective using positional coordinates. In the three-dimensional case, even the use of spherical coordinates misses some of the control aspect, as it only takes into account the turning angle in the XY plane and the angle with the vertical axis, along with step lengths. This misses the fact that, in a similar fashion to how an airplane rarely makes simply a right or left turn or a pitch up or down without changing its body orientation (rolling), the *Daphnia* behave in the same way. We capture this by taking into account the agent's behavior along its longitudinal axis.

We construct a simulation procedure based on a model constructed from information gathered in an observed swimming track from a real *D. magna*. We recognize that, to demonstrate further that our modeling approach works well, we need to observe more organisms to determine whether this behavior is consistent. However, modeling the behavior of other agents will involve a change in the parameters of the model and not necessarily require substantial changes to the model itself. This is a major gain over using positional coordinates and modeling them via ARIMA processes, as is done by Uttieri et al. (2004) and Spade et al. (2024). Our simulation procedure compares favorably to our observed track with respect to autocorrelation, distribution of step lengths and rotation angles, and smoothness of the simulated trajectory. While we do miss some autocorrelation in the step lengths, our model stacks up well otherwise against what we observe in the true trajectory.

From this point, it is possible to examine the behavior of the *Daphnia* under other settings. We have a baseline model that we can adapt to describe how the agent's behavior changes with the introduction of a predator scent or the introduction of food. We can model these changes with mixtures of similar models with different parameters, or with mixtures of different models altogether. With a grasp on the control behavior of the animal, we can account for limitations and abilities based on morphology. The model we present here provides a beginning to using control inputs as a means to describe how the behavior of the agent changes under various external stimuli. This can lead us to the ability to observe a trajectory, identify changepoints, and make inferences about what caused these changes. Furthermore, the water surrounding the agent also moves in six degrees of freedom. While this work does not account for that

component of the agent's motion because of the stillness of the surrounding water, it gives us a way to isolate how much of the *Daphnia* movement is due to its own control inputs. Having this information will facilitate modeling whilst accounting for movement of the surrounding fluid.

Declarations

Funding

None

Competing Interests

The authors declare no competing interests.

References

Ahmed, D., Benhamou, M., Bonsall, M., and Petrovskii, S. (2021). Three-dimensional random walk models of individual animal movement and their application to trap counts modelling. *Journal of Theoretical Biology*, 524:110728.

Bailey, J., Wallis, J., and Codling, E. (2018). Navigational efficiency in a biased and correlated random walk model of individual animal movement. *Ecology*, 99(1):217–223.

Benhamou, S. (2004). How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, 229(2):209–220.

Benhamou, S. (2006). Detecting an orientation component in animal paths when the preferred direction is individual-dependent. *Ecology*, 87(2):518–528.

Benhamou, S. (2019). Mean squared displacement and sinuousity of three-dimensional random search movements. *arXiv*, 1801.02435.

Benhamou, S. (2023). Of heading, posture, and body rotations derived from data acquired by animal-borne accelerometers, magnetometers, and gyrometers, kernel density estimation of the corresponding spherical distributions, and fine-scale movement reconstruction. *ArXiv*, 2310.05820.

Bovet, P. and Benhamou, S. (1988). Spatial analysis of animals' movements using a correlated random walk model. *Journal of Theoretical Biology*, 131(4):419–433.

Broadbent, S. and Kendall, D. (1953). The random walk of trichostrongylus retortaeformis. *Biometrics*, 9(4):460–466.

Byers, J. (2001). Correlated random walk equations of animal dispersal resolved by simulation. *Ecology*, 82(6):1680–1690.

Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer.

Chatterjee, N., Wofson, D., Kim, D., Velez, J., Freeman, S., Bacheler, N., Shertzer, K., Taylor, J., and Fieberg, J. (2024). Modelling individual variability in habitat selection and movement using integrated step-selection analysis. *Methods in Ecology and Evolution*, 15:1034–1047.

Codling, E., Plank, M., and Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834.

Craighead, F. and Craighead, J. (1965). Tracking grizzly bears. *BioScience*, 15(2):88–92.

Craighead, F., Craighead, J., and Davies, R. (1963). *Radiotracking of grizzly bears*. Pergamon Press.

Davenport, J. and Bels, V. (2023). Quadrupedal terrestrial locomotion in emperor penguins (aptenodytes forsteri). *Journal of Natural History*, 57(41-44):1972–1983.

Deneubourg, J., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., and Chétien, L. (1990). The dynamics of collective sorting robot-like ants and ant-like robots. *Proceedings of the International Conference on Simulation of Adaptive Behavior*, pages 356–363.

Dupont, F., Marcoux, M., Hussey, N., and Auger-Méthé, M. (2025). Improved order selection method for hidden markov models: a case study with movement data. *Methods in Ecology and Evolution*, 16(6):1215–1227.

Ephraim, Y. (2002). Hidden markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569.

Fortin, D., Morales, J., and Boyce, M. (2005). Elk winter foraging at fine scale in yellowstone national park. *Oecologica*, 145:335–343.

Garcia, R., Moss, F., Nihongi, A., Strickler, J., Göller, S., Erdmann, U., Schimansky-Geier, L., and Sokolov, I. (2007). Optimal foraging by zooplankton within patches: the case of daphnia. *Mathematical biosciences*, 207(2):165–188.

Gómez-Laich, A., Yoda, K., and Quintana, F. (2018). Insights into the foraging behavior of magellanic penguins (spheniscus magellanicus). *Waterbirds*, 41(3):332–336.

Kareiva, P. and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. *Oecologica*, 56(2–3):234–238.

Keiyu, A., Yamazaki, H., and Strickler, J. (1994). A new modelling approach for zooplankton behaviour. *Deep Sea Research Part II: Topical Studies in Oceanography*, 41(1):171–184.

Klappstein, N., Michelot, T., Fieberg, J., Pedersen, E., and Flemming, J. (2024). Step selection functions with non-linear and random effects. *Methods in Ecology and Evolution*, 15(8):1332–1346.

Klappstein, N., Thomas, L., and Michelot, T. (2023). Flexible hidden markov models for behaviour-dependent habitat selection. *Movement Ecology*, 11:30.

Komin, N., Erdmann, U., and Schimansky-Geier, L. (2004). Random walk theory applied to daphnia motion. *Fluctuation and Noise letters*, 4(01):L151–L159.

Lin, C. and Segel, L. (1974). *Mathematics applied to deterministic problems in the natural sciences*. Macmillan, New York, NY, USA.

Marsh, L. and Jones, R. (1988). The form and consequences of random walk movement models. *Journal of Theoretical Biology*, 133:113–131.

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

Morales, J., Haydon, D., Frair, J., Holsinger, K., and Fryxell, J. (2004). Extracting more out of relocation data: building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

Nicosia, A., Duchesne, T., Rivest, L.-P., and Fortin, D. (2017). A multi-state conditional logistic regression model for the analysis of animal movement. *Annals of Applied Statistics*, 11(3):1537–1560.

Okubo, A. (1980). *Diffusion and ecological problems: mathematical models*. Springer, Berlin.

Patlak, C. (1953). Random walk with persistence and external bias. *Bulletin of Mathematical Biophysics*, 15:311–338.

Pearson, K. (1905). The problem of the random walk. *Nature*, 72:294.

Pedersen, E., Miler, D., Simpson, G., and Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, 7:e6876.

Spade, D., Aliyu, I., van Horen, J., and Strickler, J. (2024). A class of statistical models for the motion of daphnia over small time scales. *Mathematical Biosciences*, 367:109114.

Straus (1820). Daphnia magna in gbif secretariat. *GBIF Backbone Taxonomy*.

Uttieri, M., Grazia-Mazzocchi, M., Nihongi, A., Ribera D'Alcalà, M., Strickler, J., and Zambianchi, E. (2004). Lagrangian description of zooplankton swimming trajectories. *Journal of plankton research*, 26(1):99–105.

Wehner, R., Marsh, A., and Wehner, S. (1992). Desert ants on a thermal tightrope. *Nature*, 357(6379):586–587.

Whoriskey, K., Auger-Méthé, M., Albertsen, C., Whoriskey, F., Binder, T., Krueguer, C., and Flemming, J. (2017). A hidden markov movement model for rapidly identifying behavioral states from animal tracks. *Ecology and Evolution*, 7(7):2112–2121.

Zucchini, W., MacDonald, I., and Langrock, R. (2016). *Hidden Markov models for time series: an introduction using R*. CRC Press, Taylor & Francis.