

SIMULATION AND ANALYSIS OF ANIMAL MOVEMENT PATHS USING NUMERUS MODEL BUILDER

Wayne M. Getz
Ludovica Luisa Vissat

Depart. ESPM, University of California
Berkeley, CA 94720-3114, USA
wgetz@berkeley.edu, l.luisavissat@berkeley.edu

Richard Salter

Numerus Inc.
850 Iron Point Rd.
Folsom, CA 95630, USA
richard.salter@numerusinc.com

ABSTRACT

We address the question of how best to fit animal movement paths, represented by point relocation time series, to a novel stochastic walk model—referred to as M-cubed—in a way that captures movement patterns at several different spatio-temporal scales. We test our approach on simulated data obtained from a high-frequency, multi-mode model constructed using the Numerus Model Builder platform. The advantage of using simulated over empirical data is that we know the processes generating movement. Fitting M-cubed to data requires that we extract movement modes and phases at various scales ranging from subhourly to daily. After fitting the M-cubed model, we evaluate how well it captures movement patterns and find that it performs exceptionally well at minute to hourly scales but needs to be extended to capture daily scale patterns driven by particular landscape features.

Keywords: random walks, multi-CAM metaFuME Markov model, cluster analysis, step length, turning angle

1 INTRODUCTION

The purpose of this paper is to present and evaluate the utility and performance of a new method for numerically simulating an animal movement path. The importance of the models arising from this new method ranges from gaining insights into the relationships between movement and disease in animal populations (Altizer, Bartel, and Han 2011) to understanding the impacts of global warming on population viability (Grémillet and Boulinier 2009). Our method is based on segmenting empirically acquired, relatively high-frequency, relocation data (i.e., minute-by-minute sequential locations of an individual as it moves over a landscape) into moderate scale “Canonical Activity Modes” (CAMs: scales ranging from tens of minutes to several hours) then extracting finer scale statistics of step-sizes, turning angles and correlations among these values for each activity mode. Following (Getz 2019), we refer to this model as an M-cubed model; which is an acronym, clarified below, for multi-CAM metaFuME Markov model.

High-frequency relocation data appropriate for the extraction of an M-cubed model is relatively expensive to come by and not yet widely shared by groups currently collecting such data for specific systems. Here, in place of high-frequency empirical data, we use relocation data generated from a multi-mode, biased, correlated, random walk (MBCRW) model. The multi-mode component is used to create specific kinds of local movement patterns (e.g., browsing within and movement among resource acquisition patches). The biases component is used to add directed movement to the global path structure. In the next section,

we define the tasks needed to construct our new M-cubed model from a set of relocation data (in this case generated by simulating our MCBRW model). We then provide an example of this workflow and compare features of the relocation data sets generated by simulating the two different models (MCBRW and extracted M-cubed). We evaluate the extent to which an M-cubed model is able to identify different local modes in a movement path, as induced by step-length and turning-angle switches among modes in the movement patterns produced by our MCBRW model. We also discuss how the M-cubed method for building an extracted model from high-frequency movement data might be enhanced to enable it to capture global structures in animal movement paths.

The standard form of a relocation data set in the x - y plane for an individual movement path that nominally begins at $t = 0$ and ends at $t = t_f$, takes the form

$$W(0, t_f) = \{ (t; x(t), y(t)) | t = 0, \dots, t_f \}. \quad (1)$$

For convenience, we selected the units of t to be the time between two consecutive relocation points. Such movement paths reflect elements of the underlying movement behavior repertoires of animals that typically are influenced by seasonally related episodic and cyclic phases. Embedded within the structure of seasonal phases we are likely to observe finer movement modes that switch amongst one other at a within-day scale. Thus movement at the daily scale can be viewed as a mixture of movement behavioral modes, also known as Canonical Activity Modes (CAMs; Getz and Saltz 2008), which themselves can be segmented into fixed-length segments of Fundamental Movement Elements (FuMEs; Getz and Saltz 2008) that are referred to as metaFuME segments. Additional factors influencing the movement pathway structure of individuals may include an individual's physiological or emotional state, the time-of-day, territoriality, distances from attractive centers (e.g., watering points, nests), the presence of both conspecifics and heterospecific competitors and enemies, the physical and ecological structure of the landscape itself, and memories of favorable locations both near and far that provide access to food, water, and safety.

Our ability to extract the role that the various factors listed above play in determining the mode and phase structure of $W(0, t_f)$ (Eq. 1) depends on the resolution of the data—specifically, the frequency at which it was collected or the units of t . Coarse data (e.g., hourly or subdaily) may only permit us to uncover environmental factors driving seasonal movement phases, intermediate resolution data (e.g., subhourly) may only allow us to uncover environmental and internal state factors driving diel (daily) cycles, while high resolution data (minute or sub-minute) is needed to answer questions that relate to behavior lasting a few to several minutes.

No matter the resolution of the relocation data, however, the first step in dealing with such data is to extract bivariate step-length (SL, $s(t)$) and turning-angle (TA, $a(t)$) time-series, using the approach described below. Once this has been done, then various methods have been used to extract information from the derived bivariate time series ($s(t), a(t)$), along with available covariate data (environmental, individual state, location of conspecifics, and also of heterospecific competitors and enemies), to confront the challenge mentioned above. The efficacy of such methods—which include hidden Markov models (Zucchini et al. 2016) and behavioral change point analysis (Gurarie et al. 2009, Gurarie et al. 2016)—is most easily evaluated using data that is derived from processes known to be causative in driving the observed movement patterns. One way to do this is to analyse simulation data derived from movement models of varying complexity (e.g., MCBRW models) to see how much information can be extract from the relocation and covariate data sets.

In summary, relocation data sets of the type depicted in Eq. 1 can either be empirical, (which we flag using the notation $W^e(0, t_f)$) or derived from simulation models (which we flag using the $W^s(0, t_f)$). In addition, the simulation model can be fitted to empirical relocation data using some appropriately formulated “best-fitting” procedure.

We use the notation $W_i^s(0, t_f)$ to denote the i^{th} instance of the relocation data set derived from the i^{th} simulation of a model that “best fits” an empirical relocation data set $W^e(0, t_f)$. In addition, we use calligraphic notation to denote sets of sets, so that an ensemble of N simulated “best fitting” walks W_i^s is denoted by

$$\mathcal{W}(0, t_f) = \{W_i^s(0, t_f) | i = 1, \dots, N\} \text{ is a “best-fitting” ensemble to empirical path } W^e(0, t_f).$$

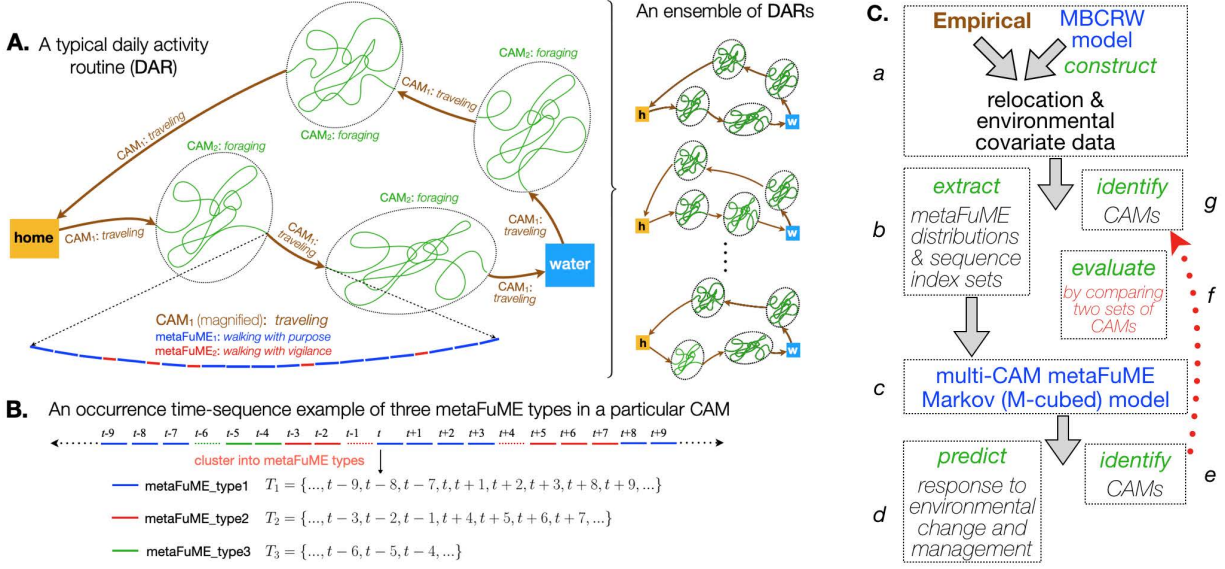


Figure 1: Segmentation of a daily activity routine (DAR) of a central-placed forager. The logic behind our study. **A.** Left-hand side: a simple two-CAM DAR with water stop is depicted along with a magnified depiction of a *traveling* CAM generated from two types of metaFuMEs. Right-hand side: an ensemble of DARs of a particular type (h=home, w=water). **B.** An illustration of the time sequence index sets T_i , $i = 1, 2, \dots$, obtained from an analysis of the metaFuME types that constitute a particular CAM sequence. (Figure adapted from Getz 2019). **C.** The logic of fitting the MBCRW model to data to for the production of ensembles of DAR segments that can then be analysed as surrogates for real data (see text for discussion). Model names are in blue and actions are in green.

The challenge lies in first finding a suitable class of appropriate complexity models with which to model the process producing the data $W^e(0, t_f)$. It then moves on to finding the best-fitting set of parameters for this model such that the data $W^e(0, t_f)$ are more likely to lie within the ensemble $\mathcal{W}(0, t_f)$ than any other ensemble that can be generated by the model using a different set of parameters. Solving the above problem represents a program of research and study that will be with us for decades to come because it is both model-structure and estimation-method dependent. Models with improved and refined structures will be proposed over time as we learn more about the process that generate the movement paths of individuals belonging to particular species. In addition, we will steadily develop better methods of estimation as computational technologies and parameter-fitting algorithms improve over time.

The approach we take draws upon a hierarchical, appropriate complexity, segmentation framework—using methods recently reviewed in Getz 2019—that we use to breakdown empirical Life-history Tracks (LiTTs) into particular Life-history Movement Phases (LiMPs) that then consist of a series of Diel Activity Routines (DARs), which characterize each LiMP (e.g., a ranging phase versus a migratory phase—see Getz 2019 for further discussion). These DARs can then be further segmented into sequences of different types of canonical activity modes (CAMs; Fig. 1A.) that, in turn, can be modeled as sequences of “meta” fundamental

movement elements (metaFuMEs). Additionally, we assume that each type of CAM is constructed from the same basis set of metaFuMEs, with differences among CAM types characterized by the lengths and frequencies in the occurrence of particular metaFuME sequences (Fig. 1B.). As an aside, we note that fundamental movement elements themselves, as discussed in (Getz and Saltz 2008, Getz 2019), cannot generally be extracted directly from relocation data. Rather, current data are typically only suitable for identifying CAMs from step-size and turning-angle time series using hidden Markov model (HMM) and behavioral change point analysis (BCPA) methods. Such methods have been used to identify CAMs at a particular sampling frequency resolution. Aggregating all the CAMs of the same type together allows us to construct bivariate step-length and turning-angle distributions for the CAM ensembles.

2 GOALS, TASKS AND WORKFLOW

For clarity, we provide an outline of the tasks undertaken in our study and the workflow needed to carry out the required analyses and simulations. Our methods are intended to be applied to relatively high-frequency empirical data (viz., collected at frequencies around 1 to 0.01 Hz). Lacking such data, and also taking up the challenging of building models that can be fitted to such data for predictive purposes (e.g., addressing management and response to global change questions), we first focus on building stochastic simulation models that can realistically simulate daily activity routines (DARs) with respect to their underlying CAM structure. We refer to this model as our multi-mode, biased, correlated random walk (MBCRW) model.

As shown in Figure 1C., we use our MBCRW model to produce ensembles of DAR segments that can then be analysed as surrogates for real data (a). The approach we take is to extract a best-fitting basis set of metaFuMEs and associated sequence index sets (b) that can then be used to construct a multi-CAM metaFuME Markov (M-cubed) model (c). This M-cubed model, extracted from our surrogate data, could also be extracted directly from empirical data, collected at the appropriate frequency. When extracted either from surrogate or real data, the M-cubed model can be used to predict responses to changes in environment and management (d) but also CAMs can be identified (e) to see how well the M-cubed model reproduces (f) the CAMs evident in the original empirical/surrogate data (g).

We use a stochastic process model to generate relocation data. These relocation data are created at a frequency of 0.2 Hz (i.e., locations are generated five seconds apart) for a 24 hour period (i.e., 17,280 time steps). Thus we use the model to create ensembles of DARs. The output from this model, evenly subsampled to create data sets with an order of magnitude fewer points will be used as surrogate empirical DAR segments to create ensembles $W^s(0, 24h)$ (Eq. 1) of relocation data sets. These data will then be used to apply our deconstructive methods and use the results to evaluate our approach. In particular, the relocation sets that we use to represent these DARs will be a subsampling of every sixth point of the generated data. This provides us with relocation sets for which one unit of t equals 30s, implying $t_f = 2,880$ units over a 24 hour period. The tasks involved are the following:

A. Simulation of DAR ensembles

- A1. We provide a mathematical description of the underlying simulation model as a generalized random walk with step length and directional heading correlations and biases that depend on the time of day, structure of local environment, and knowledge of the regional location of critical resources.
- A2. We build the model using the Numerus Model Builder software platform and generate the code necessary to run our simulations.
- A3. For a selected set of model parameter values, we run the model to obtain an ensemble of N DARs.

B. Deconstruction of DAR ensembles

- B1. We prepare our simulated relocation data sets for analysis by thinning them from 17,280 to 2,880 points for each DAR, excluding the initial location (0,0).

- B2.** We carry out BCPA with running windows of different widths (e.g., 5, 20 and 60) or HMM analysis to identify both short and long duration CAMs.
- B3.** From the set of CAMs, we extract step length (SL) and turning angle (TA) time series and calculated means, variances, auto-correlations and covariances (cross-correlations).
- B4.** We perform cluster analysis on SL and TA point locations statistics to extract metaFuME sets and accompanying metaFuME sequence index sets.
- B5.** We group SL and TA data identified in B4 as belonging to metaFuME of a particular type, and we use the data in the sequence index sets to estimate metaFuME transition probabilities. We also estimate transition probabilities between CAMs using the CAM sequence extracted in B2.

Once A. and B. are completed, we use the metaFuME and CAM data sets and transition probabilities estimated in B5 to simulate ensembles of walks. We then compare these ensembles to those generated under Task A to assess similarities and differences. In particular, in our study we will look at SL and TA distributions, time spent in each CAM, area coverage and distance from a given point to compare the two ensembles.

3 MODEL: DESCRIPTION AND IMPLEMENTATION

In terms of movement models, two contrasting approaches are common. The first is the use of numerical event-oriented (e.g. Gillespie's event oriented algorithm—(Gillespie 1976, Gillespie 1977), or time-discretizations methods to compute solutions to stochastic differential equations that model continuous time stochastic processes. In increasing order of complexity, these include a pure random walk (Wiener process; e.g. see (McClintock et al. 2014)), a linear- and angular-velocity-biased random walk (Ornstein-Uhlenbeck process, (Gurarie et al. 2017)), and location salient random walk (e.g., walks in a force field, (Magdziarz et al. 2012); or movement dependent on the local landscape, (Harris and Blackwell 2013)), where the location itself could be a function of time (e.g., the centroid of a moving group—see (Langrock et al. 2014)). The second is to use rule-based simulations of movement over rectangularly or hexagonally rasterized landscapes without the benefit of a compact underlying differential equation structure (Getz et al. 2015, del Mar Delgado et al. 2018)

Our path simulation algorithm of the movement of individuals over landscapes is an outgrowth of Brownian motion (Pozdnyakov et al. 2014), correlated random walks (Kareiva and Shigesada 1983) and, even Levy walks (Benhamou 2007) simulation methods. Within the context of appropriate modeling methods (Getz et al. 2018), our elaborations to these earlier approaches in formulating our MBCRW include both biased directionality and mixed-distribution (Morales et al. 2004) components. The multiple modality aspect of our model arises in two ways. First, from switching movement modes locally at a frequency that assumes a particular level of landscape heterogeneity (e.g., moving within and among resource acquisition patches). Second, from switching directional movement biases at some slower daily-activity-related frequency in terms of attraction to distant (peripheral) geographic locations such as water holes or resource rich areas, and then attraction back to centrally located nest sites or home range centers. A detailed mathematical description of this model can be found in the supporting online file ([SOF](#)). In short, our model is an MBCRW, with daily walks between and around two points of attraction, that always starts around the same central point, followed by a daily excursion towards and around a peripheral point that may change each day, with a final end-of-each-day movement back towards the central point (see right-hand panel in Fig. 2).

Our model is implemented using the Numerus Model Builder (NMB) platform, which allows us to code a digital simulation using a set of graphically implemented, drag-and-drop icons, as shown in Figure 2, in which mathematical expressions of model equations are inserted. NMB was chosen since it facilitates a quick and accurate representation of the model and greatly diminishes coding tasks, as discussed in more detail in (Getz et al. 2015, Getz et al. 2018). Once all the necessary equations and segments of code

have been inserted into equation windows and code chip frames, NMB then generates a script that can be used to run simulations of the model over the desired interval of time. The NMB modelling panel and the output of the simulation for the MBCRW model are shown in Fig. 2. We generate the set of relocation data $W_{[0,T]}^s$ by simulating the model for a 10-day temporal horizon ($T = 172,800$), where one time-unit in NMB corresponds to 5s and one spatial unit to 100m. In [SOF](#) we provide the parameters used in this simulation.

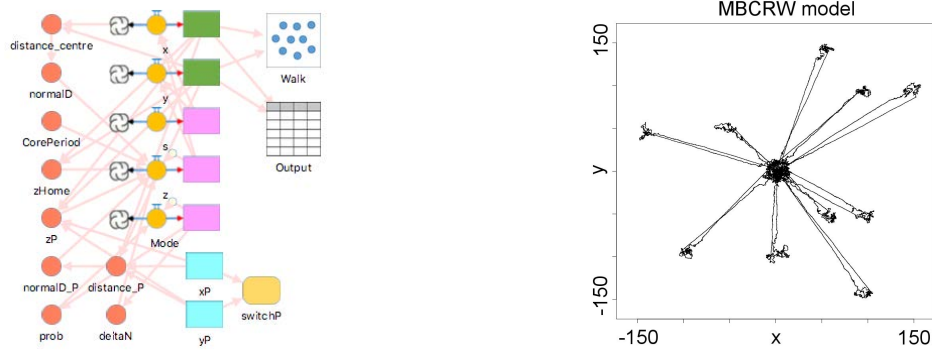


Figure 2: A graphical construction of our NMB model (left panel; values and equations are inserted into a form that is opened when a component icon is selected) and output from one instance of a stochastic 10-day simulation (right panel)

4 DECONSTRUCTION OF DAR ENSEMBLES

The approach we develop here is to use the time interval that corresponds to the same frequency at which the data were collected and to either use actual step-size and turning-angle data, or smoothed/idealized functions that are fitted to or theoretically represent these data. As part of the simulation, considerations of directional biases, step-size and turning-angle serial correlations, step-size/turning-angle cross-correlations, and context-dependent environmental and individual internal-state step-size and turning-angle distributions may be considered (Ahearn et al. 2017, Langrock et al. 2014). Although, individuals may also take the locations and directions of heading of other individuals into account as they move across landscapes (Couzin et al. 2005, Conradt et al. 2009, Langrock et al. 2014), we do not consider this level of complexity here.

We undertook analyses of the relocation data using the R Studio implementation of the R programming language, where in our case these data were obtained in a CSV format from our NMB simulations. Our analyses were designed to extract from the data all the information needed to construct our M-cubed model.

First, we subsampled the output of the NMB simulation data at a frequency appropriate for extracting relatively fast canonical activity modes (i.e., CAMs at a sub-hourly to hourly scale—see Getz 2019) by selecting every 60th point, which corresponds to a 5 min interval between relocation points and is in the range of relocation data that is now practical to collect over multiple months. Using the R package `moveHMM`, we then applied the Viterbi algorithm to sort our subsampled relocation data into the most likely sequence of CAMs and identified every distinct CAM as one of K types. Each CAM segment can be represented by

$$W_{[t_i+1, t_{i+1}]}^s = \{(x_t, y_t) \mid t = t_i + 1, \dots, t_{i+1}\},$$

such that a complete path $W_{[0,T]}^s$ is then made up of p such segments strung together over an interval $[0, T]$: i.e.,

$$W_{[0,T]}^s = \{W_{[0, t_1]}^s W_{[t_1+1, t_2]}^s W_{[t_2+1, t_3]}^s \cdots W_{[t_{p-1}+1, T]}^s\},$$

with a corresponding set of $p - 1$ change points $T_{CP} = \{t_1, t_2, \dots, t_{p-1}\}$.

Second, we subsampled the output of the NMB simulation data at a frequency appropriate for extracting metaFuMEs (i.e., around 10 s to 1 min—see Getz 2019) by selecting every 5th point, which corresponds to a 30 s interval between relocation points. Thus, given the extracted sequence of CAM at a 5-min relocation point interval, we assigned the same CAM index to each of the corresponding 10 time-points included in each step of the 5-min interval identified in the first step as belonging to a particular CAM type.

For each resulting segment at a 30 s time-scale, we extracted the step length $s(t)$ and turning angle $a(t)$ time-series, by calculating the following values:

$$\begin{aligned} s(t) &= \sqrt{(x(t+1) - x(t))^2 + (y(t+1) - y(t))^2}, \\ a(t) &= \arctan \frac{y(t+1) - y(t)}{x(t+1) - x(t)} - \arctan \frac{y(t) - y(t-1)}{x(t) - x(t-1)}, \end{aligned}$$

for $t = 1, \dots, T$. Note that for $t = 0$ $a(t)$ is not defined. We then obtained the mean and the variance (μ_x, σ_x) , where $x = s, a$, and calculated the running-term auto-correlation time-series $v^{xx}(t)$ for $t = 2, \dots, T$ and running-term cross-correlation time-series $v^{as}(t)$ for $t = 1, \dots, T$:

$$v^{xx}(t) = \frac{(x(t) - \mu_x)(x(t-1) - \mu_x)}{\sigma_x}, \quad v^{as}(t) = \frac{(a(t) - \mu_a)(s(t) - \mu_s)}{\sqrt{\sigma_a \sigma_s}}.$$

We also computed the time-series of the turning-angle cosines, which we subsequently used in our metaFuME cluster analysis in place of the the turning angles themselves: i.e., we generated the values:

$$c_a(t) = \cos(a(t)).$$

We performed data standardization through min-max normalization (Van Moorter, Visscher, Jerde, Frair, and Merrill 2010), as data pre-processing before carrying out a cluster analysis:

$$\bar{x}_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}.$$

We collected the standardized sets $\bar{s}(t)$, $\bar{c}_a(t)$, $\bar{v}^{ss}(t)$, $\bar{v}^{c_a c_a}(t)$ and $\bar{v}^{as}(t)$ for each t identified as part of the same CAM to perform cluster analysis. We used Ward's method, provided by the R package `stats`, to extract the set of clusters of metaFuMEs for each CAM and the elbow method to estimate the optimal number of clusters. Once we have extracted the clusters of metaFuMEs for each CAM, we evaluated if some metaFuME clusters could be associated and grouped together. To accomplish this, we performed cluster analysis of mean and variance of $\bar{s}(t)$ and $\bar{c}_a(t)$ values for each metaFuME initial cluster, again using Ward's method and then estimating the optimal number of clusters through the elbow method. The results of the elbow method are shown in the [SOF](#) (Figure S2). Once we have identified the final clusters of J metaFuMEs, we extract the corresponding sets of step-lengths S_j and turning-angles A_j for each metaFuME type $j = 1, \dots, J$. In our study, when simulating the M-cubed model, we sampled from these data, according to the chosen or current metaFuME type such that each element in S_j and A_j had an equal probability of being selected.

As a following step in our analysis, we extracted the sequences of metaFuME types for each CAM segment. Using the R package `markovchain`, we fitted the transition probability matrix for each sequence and calculated the weighted average (depending on the sequence length) to extract the metaFuME transition probability matrices for each CAM. Given n sequences for a given CAM, the weighted average for a matrix entry was calculated as follows:

$$\bar{m}_{ij} = \frac{\sum_{s=1}^n m_{ij}^s \times l_s}{\sum_{s=1}^n l_s},$$

where l_s is the length of the sequence s and m_{ij}^s is the (i, j) entry of the matrix extracted from sequence s . Note that some CAMs might not contain all the J metaFuMEs. Assuming that M_{S_k} indicates the index set of the metaFuME types that compose a given CAM k , for $i \notin M_{S_k}$ the matrix M_k will need to contain entries m_{ij} equal to f_j^k , for $j \in M_{S_k}$, and equal to 0, for $j \notin M_{S_k}$, where f_j^k represents the frequencies of metaFuME j in CAM k . Moreover, the entries m_{ij} , for all i and $j \notin M_{S_k}$, will be equal to 0. Defining these rows in the transition probability matrices, facilitates the change to the correct metaFuME sets when switching between CAMs.

Recalling the first task of our analysis, we identify the most likely sequence of CAMs again using the Viterbi algorithm. In this case, however, we used the R package `markovchain` and the CAM sequence to extract the CAM transition probability matrix M_C . In our simulation, we assumed the same CAM always persisted for at least 10 time points (i.e., 5 mins) and therefore we evaluated the transitions between CAMs using M_C every 10 time points.

5 M-CUBED MODEL SIMULATION

By performing the tasks defined in Section 2, and described in more detail in Section 4, we extracted all the data related to metaFuMEs and CAMs which are necessary to construct an M-cubed model. To summarise, these data are the following: a finite set of J metaFuME SL and TA variables set S_j and A_j , a $K \times K$ transition probability matrix M_C , representing the probability to switch between CAMs, and K , $J \times J$ transition probability matrices M_k , $k = 1, \dots, K$ for each of the K CAM types, to describe the transition probabilities between metaFuMEs while in a given CAM type.

The simulation of this model produces a set of relocation data $W^s(0, T_M)$, as described in Algorithm 1 (Fig. 5A.). This algorithm is an amalgam of an “outer” and “inner” process. The outer process is a Markov transition process governing the probability that the current CAM will switch to a different CAM at every 5-min time interval. The inner process is a within-CAM stochastic dynamical systems process that controls the change of location from one time-step to the next (30 s time-scale) according to the selected metaFuME.

Algorithm To initialize the simulation, we declare the initial location (x_0, y_0) , initial CAM k , initial metaFuME j , initial heading θ_0 and final time T_M . For each time step t , we generate location data by drawing step length and turning angle $(s_t^j$ and $a_t^j)$ pairs from the sets S_j and A_j , according to the metaFuME type j in operation at time t . The location data and the elapsed times are updated every step and the switch between metaFuME types is ruled by the matrix M_k (when in CAM k). After the given time in CAM k has elapsed (10 time steps, counted using the variable \bar{t}), the CAM update is defined by the transition probability matrix M_C . We use the function f to represent the process by which both metaFuME and CAM values are updated according to their current values and the given transition probability matrices.

6 M-CUBED MODEL PARAMETERS

In our study, the extracted M-cubed model has the following features. It is composed of 3 CAM types and 3 metaFuME types. Results of the HMM analysis are shown in Fig. 3.

From Fig. 3, we see that CAM 1 represents random movement mostly around the centers of attractions while CAM 2 represents more correlated movement around these centers. On the other hand, CAM 3 represents highly directed movement that occurs between centers of attraction. From our cluster analysis (details in [SOF](#)) CAM 1 is composed of metaFuMEs of type 1 and 2 (frequencies: 0.51, 0.49), CAM 2 of metaFuMEs of type 1, 2 and 3 (frequencies: 0.4, 0.1, 0.5) while CAM 3 by type 1 and 3 (frequencies: 0.16, 0.84). Distributions of SL and TA for the 3 metaFuME types are shown in Fig. 4. MetaFuME of type 1 is a *variable step/persistent direction* metaFuME while metaFuME of type 2 is a *variable step/random direction* one. MetaFuME 3 is a *large step/persistent direction* metaFuME type.

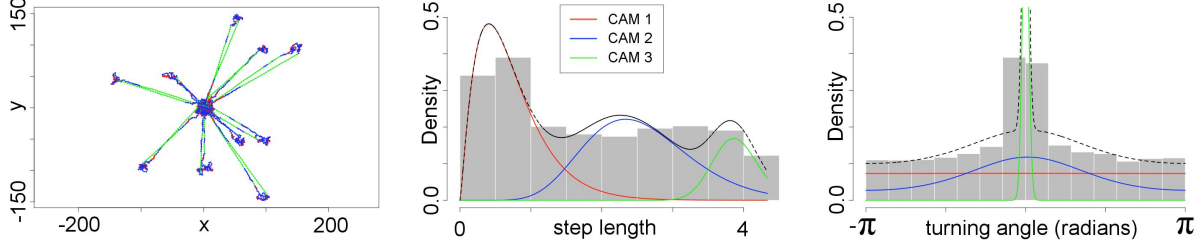


Figure 3: Results of the CAM extraction from our HMM analysis (plotted using the R package `moveHMM`). Left panel: CAM segments plotted using legend colors. Middle panel: SL distribution in 100m bins. Right panel: TA distribution in radians.

The CAM transition probability matrix M_C and the matrices M_k for each CAM of type $k = 1, \dots, K$ for the M-cubed analysis of the MCBRW data generated for this study can be found in our [SOF](#). The initial configuration of our M-cubed model was (0,0) for the initial location, CAM 1 as initial CAM and metaFuME 1 as initial metaFuME. We implemented this model in NBM and run a simulation for 28,800 time steps (10 days). The NMB modelling panel and the output of the simulation are shown in Fig. 5B,C. Additional examples of the M-cubed model simulation output are provided in our [SOF](#) (Figure S3).

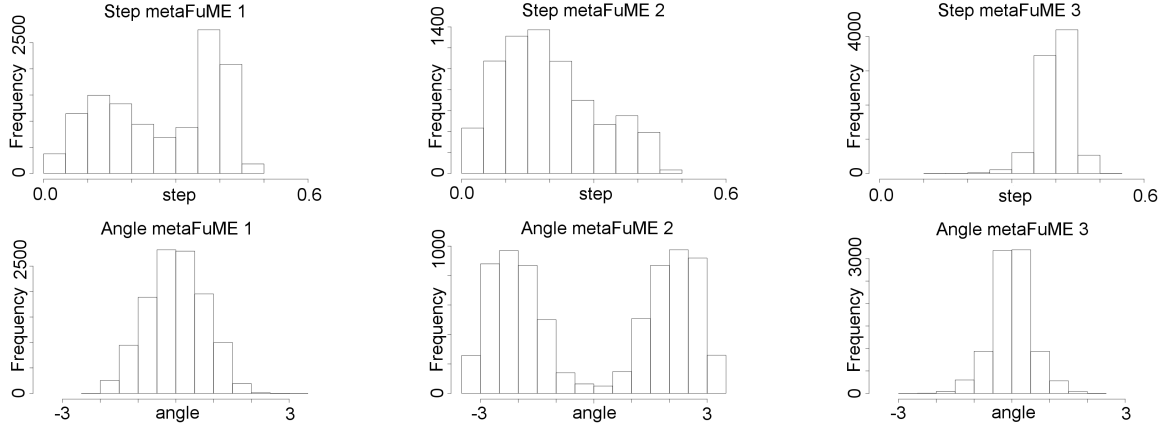


Figure 4: Distribution of SL (100m bins) and TA (radians) for metaFuME types, from subsampled MCBRW data

7 MODEL COMPARISONS

The structure of the paths obtained from our MCBRW and M-cubed models are very similar at the metaFuME and multi-CAM scales by design. This is illustrated in [SOF](#) where the SL and TA distributions are plotted side by side in Fig. S4 and S5 for each of the CAM types. We used the R package `trajr` to calculate the sinuosity of the two simulated paths and, unsurprisingly, obtained similar results because sinuosity is a relatively fine scale structure of a movement path: viz., 2.75 for the MCBRW model and 2.90 for the M-cubed model. Moreover, we extracted the time spent in each CAM, for both models (Fig. S6 of [SOF](#)) and these were similar as well. The two models, as expected, varied quite considerably, when comparing the more global scale features of the paths they produce. In particular, the total area covered over time, the area covered each day and the distance from the origin over time (Fig. S7, S8 and S9 in [SOF](#)) were notably different. These results confirm that the M-cubed model captures local features of movement paths (e.g., SL and TA distributions and sinuosity) but fails to capture global features which arise in our MCBRW model from movement bias due to the presence of centers of attraction that periodically switch over time.

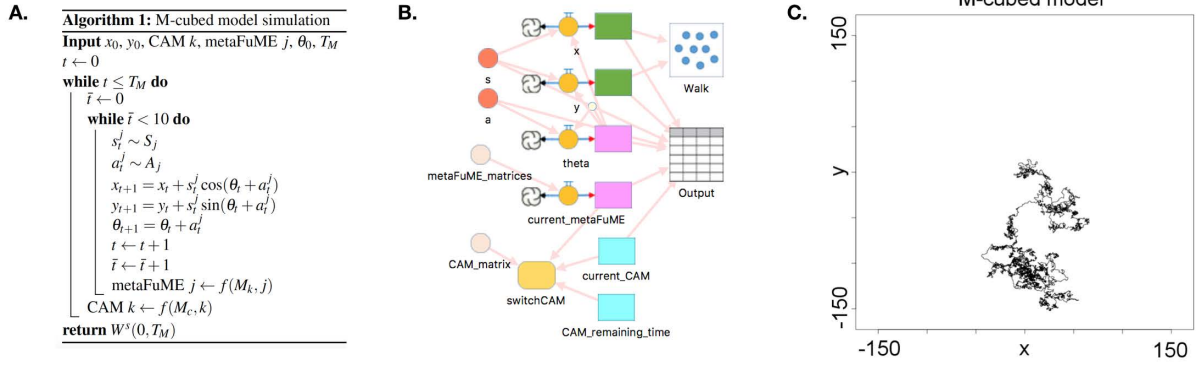


Figure 5: **A.** Algorithm 1. **B.** NMB model canvas. **C.** One instance of a simulation of our M-cubed model

Thus, derived statistical movement models, such as M-cubed, need to be improved to capture global features by considering not only turning angles, but actual angles of heading.

8 CONCLUSION

Recently, stochastic walk models that do not include switches in directional movement bias beyond metaFuME and CAM time scales, have been promoted as a way to capture global features of walks from empirical relocation data—e.g., the use of autocorrelated-kernel-density estimation (AKDE) to compute home range size (Fleming and Calabrese 2017, Noonan et al. 2019) from movement paths. Such methods, however, as demonstrated through our comparison of output from our MBCRW and M-cubed models, will fail to provide reliable estimates of global properties of walks when the relocation data is fitted to models that fail to take directional biases into account at the multi-CAM scale. This includes the Ornstein-Uhlenbeck stochastic process model that is the bases for path analysis and home range estimation using the increasingly popular `ctmm` R package (Calabrese, Fleming, and Gurarie 2016). As with the Ornstein-Uhlenbeck model, our M-cubed models fails to take account for key processes that primarily affect global rather than local structure of paths and thus needs to be extended to include such processes. The answer is not to try to fit our MBCRW model directly to relocation data since, as with any approach that tries to simultaneously fit data created by processes occurring at different scales, the approach is likely to dramatically fail. Rather, an extended M-cubed that first fits an M-cubed model to relocation data, as outlined here, and then accounts for processes taking place at the multi-CAM time scale and beyond, is needed; and is a topic that remains to be addressed in future studies, as does an evaluation of the basic M-cubed model itself using real rather than manufactured data.

ACKNOWLEDGMENTS

The development of NOVA, a precursor to Numerus Model Builder, was supported by NSF grant CNS-0939153 to Oberlin College (PI: RS) and NSF-EEID grant 1617982 (PI: WMG). The supporting online file (SOF), containing details of the MBCRW model, simulation parameter values, and additional results is available at https://ludovicalv.github.io/PDFs/SOF_SCS_2020.pdf.

REFERENCES

Ahearn, S. C., S. Dodge, A. Simcharoen, G. Xavier, and J. L. Smith. 2017. “A context-sensitive correlated random walk: a new simulation model for movement”. *International Journal of Geographical Information Science* vol. 31 (5), pp. 867–883.

- Altizer, S., R. Bartel, and B. A. Han. 2011. "Animal migration and infectious disease risk". *science* vol. 331 (6015), pp. 296–302.
- Benhamou, S. 2007. "How Many Animals Really Do the Lévy Walk?". *Ecology* vol. 88 (8), pp. 1962–1969.
- Calabrese, J. M., C. H. Fleming, and E. Gurarie. 2016. "ctmm: an R package for analyzing animal relocation data as a continuous-time stochastic process". *Methods in Ecology and Evolution* vol. 7 (9), pp. 1124–1132.
- Conradt, L., J. Krause, I. D. Couzin, and T. J. Roper. 2009. "“Leading according to need” in self-organizing groups". *The American Naturalist* vol. 173 (3), pp. 304–312.
- Couzin, I. D., J. Krause, N. R. Franks, and S. A. Levin. 2005. "Effective leadership and decision-making in animal groups on the move". *Nature* vol. 433 (7025), pp. 513.
- del Mar Delgado, M., M. Miranda, S. J. Alvarez, E. Gurarie, W. F. Fagan, V. Penteriani, A. di Virgilio, and J. M. Morales. 2018. "The importance of individual variation in the dynamics of animal collective movements". *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 373 (1746). pii. 20170008.
- Fleming, C. H., and J. M. Calabrese. 2017. "A new kernel density estimator for accurate home-range and species-range area estimation". *Methods in Ecology and Evolution* vol. 8 (5), pp. 571–579.
- Getz, W. M. 2019. "A hierarchical framework for segmenting movement paths". *bioRxiv*.
- Getz, W. M., C. R. Marshall, C. J. Carlson, L. Giuggioli, S. J. Ryan, S. S. Romanach, C. Boettiger, S. D. Chamberlain, L. Larsen, P. D’Odorico, and D. O’Sullivan. 2018. "Making ecological models adequate". *Ecology Letters* vol. 21 (2), pp. 153–166.
- Getz, W. M., R. Salter, A. J. Lyons, and N. Sippl-Swezey. 2015. "Panmictic and clonal evolution on a single patchy resource produces polymorphic foraging guilds". *PloS one* vol. 10 (8). pii. e0133732.
- Getz, W. M., R. Salter, O. Muellerklein, H. S. Yoon, and K. Tallam. 2018. "Modeling epidemics: A primer and Numerus Model Builder implementation". *Epidemics*.
- Getz, W. M., and D. Saltz. 2008. "A framework for generating and analyzing movement paths on ecological landscapes". *Proceedings of the National Academy of Sciences* vol. 105 (49), pp. 19066–19071.
- Gillespie, D. T. 1976. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *Journal of computational physics* vol. 22 (4), pp. 403–434.
- Gillespie, D. T. 1977. "Exact stochastic simulation of coupled chemical reactions". *The journal of physical chemistry* vol. 81 (25), pp. 2340–2361.
- Grémillet, D., and T. Boulinier. 2009. "Spatial ecology and conservation of seabirds facing global climate change: a review". *Marine Ecology Progress Series* vol. 391, pp. 121–137.
- Gurarie, E., R. D. Andrews, and K. L. Laidre. 2009. "A novel method for identifying behavioural changes in animal movement data". *Ecology letters* vol. 12 (5), pp. 395–408.
- Gurarie, E., C. Bracis, M. Delgado, T. D. Meckley, I. Kojola, and C. M. Wagner. 2016. "What is the animal doing? Tools for exploring behavioural structure in animal movements". *Journal of Animal Ecology* vol. 85 (1), pp. 69–84.
- Gurarie, E., C. H. Fleming, W. F. Fagan, K. L. Laidre, J. Hernández-Pliego, and O. Ovaskainen. 2017. "Correlated velocity models as a fundamental unit of animal movement: synthesis and applications". *Movement ecology* vol. 5 (1), pp. 13.
- Harris, K. J., and P. G. Blackwell. 2013. "Flexible continuous-time modelling for heterogeneous animal movement". *Ecological Modelling* vol. 255, pp. 29–37.

- Kareiva, P., and N. Shigesada. 1983. "Analyzing insect movement as a correlated random walk". *Oecologia* vol. 56 (2-3), pp. 234–238.
- Langrock, R., J. G. C. Hopcraft, P. G. Blackwell, V. Goodall, R. King, M. Niu, T. A. Patterson, M. W. Pedersen, A. Skarin, and R. S. Schick. 2014. "Modelling group dynamic animal movement". *Methods in Ecology and Evolution* vol. 5 (2), pp. 190–199.
- Magdziarz, M., R. Metzler, W. Szczotka, and P. Zebrowski. 2012. "Correlated continuous-time random walks in external force fields". *Physical Review E* vol. 85 (5). pii. 051103.
- McClintock, B. T., D. S. Johnson, M. B. Hooten, J. M. Ver Hoef, and J. M. Morales. 2014. "When to be discrete: the importance of time formulation in understanding animal movement". *Movement ecology* vol. 2 (1), pp. 21.
- 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* vol. 85 (9), pp. 2436–2445.
- Noonan, M. J., M. A. Tucker, C. H. Fleming, T. S. Akre, S. C. Alberts, A. H. Ali, J. Altmann, P. C. Antunes, J. L. Belant, D. Beyer et al. 2019. "A comprehensive analysis of autocorrelation and bias in home range estimation". *Ecological Monographs* vol. 89 (2). pii. e01344.
- Pozdnyakov, V., T. Meyer, Y.-B. Wang, and J. Yan. 2014. "On modeling animal movements using Brownian motion with measurement error". *Ecology* vol. 95 (2), pp. 247–253.
- Van Moorter, B., D. R. Visscher, C. L. Jerde, J. L. Frair, and E. H. Merrill. 2010. "Identifying Movement States From Location Data Using Cluster Analysis". *The Journal of Wildlife Management* vol. 74 (3), pp. 588–594.
- Zucchini, W., I. L. MacDonald, and R. Langrock. 2016. *Hidden Markov models for time series: an introduction using R*. Chapman and Hall/CRC.

AUTHOR BIOGRAPHIES

WAYNE M. GETZ is a Professor in the Graduate Division at the University of California, Berkeley, and a Founding Partner of Numerus Inc. His email address is wgetz@berkeley.edu.

LUDOVICA LUISA VISSAT is a Postdoctoral Scholar at the Department of Environmental Science, Policy and Management at the University of California, Berkeley. Her email is l.luisavissat@berkeley.edu.

RICHARD SALTER is an Emeritus Professor of Computer Science at Oberlin College and a Founding Partner of Numerus Inc. His email address is richard.salter@numerusinc.com.