Detecting Movement Patterns using Brownian Bridges

Kevin Buchin
Dep. of Mathematics and Computer Science
TU Eindhoven
The Netherlands
k.a.buchin@tue.nl

Stef Sijben
Dep. of Mathematics and Computer Science
TU Eindhoven
The Netherlands
s.m.a.sijben@student.tue.nl

T. Jean Marie Arseneau
Anthropological Institute & Museum
University of Zurich
Switzerland
thelmajeanmarie.arseneau@uzh.ch

Erik P. Willems
Anthropological Institute & Museum
University of Zurich
Switzerland
e.willems@aim.uzh.ch

ABSTRACT
In trajectory data a low sampling rate leads to high uncertainty in between sampling points, which needs to be taken into account in the analysis of such data. However, current algorithms for movement analysis ignore this uncertainty and assume linear movement between sample points. In this paper we develop a framework for movement analysis using the Brownian bridge movement model (BBMM), that is, a model that assumes random movement between sample points. Many movement patterns are composed from basic building blocks, like distance, speed or direction. We efficiently compute their distribution over space and time in the BBMM using parallel graphics hardware. We demonstrate our framework by computing patterns like encounter, avoidance/attraction, regular visits, and following. Our motivation to study the BBMM stems from the rapidly expanding research paradigm of movement ecology. To this end, we provide an interface to our framework in R, an environment widely used within the natural sciences for statistical computing and modeling, and present a study on the simultaneous movement of groups of wild and free-ranging primates.

Categories and Subject Descriptors
F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems—Geometrical problems and computations

General Terms
Algorithms, Theory

*Research supported by COST (European Cooperation in Science and Technology) ICT Action IC0903 MOVE, the Swiss National Science Foundation (Sinergia 133040), and the Forschungskredit of the University of Zurich.

Keywords
trajectory, geometric algorithms, movement ecology

1. INTRODUCTION
Widespread use of location-aware devices results in an increased demand for the analysis of movement data. Recent years have seen considerable progress in methods for analyzing movement data. However, most of these methods have been developed for the case of very densely sampled data for which one can assume that movement between sample points is linear, or for which it is not even necessary to interpolate between sample points. These are viable models when the sampling rate is high, but their validity quickly decreases with lower sampling rates.

Low sampling rates are however quite common for mobile location-aware devices. This is due to the trade-offs that often will have to be made between the desired sampling interval for the research question (or mobile application) on the one hand, and practical limitations posed by technological and financial means, as well as ethical considerations on the other hand. This trade-off is particularly restrictive in the newly emerging research paradigm of movement ecology, in which movement data are collected on wild and free-ranging animals, whose biology and well-being put severe restraints on e.g. the weight and size of devices. The invasiveness and difficulties of equipping a wild animal with a location-aware device (often in the form of a telemetric collar placed around the neck) imply that the attainable observation period is set by battery-life, which is approximately anti-proportional to

![Figure 1: Brownian bridge movement model. (a) Probability density between two sample points [18]. (b) Two trajectories (blue/violet) with encounter (green/red; yellow in overlap).](image-url)
the sampling rate. As a consequence the sampling rate is typically chosen below a threshold where a linear movement model would make sense. Instead models assuming random movement are used. The Brownian bridge movement model (BBMM) is a commonly used movement model assuming Brownian motion conditioned under the locations at the sample points. Figure 1 illustrates this model. Figure 1(a) shows for two consecutive sample points the probability density for the fraction of time spent in different regions (from [18]). The two peaks in the probability density correspond to the locations of the two sample points. Figure 1(b) shows the locations of two daily trajectories of wild primates with one sample point per hour. This figure also illustrates the encounter pattern. The trajectories are colored blue when the other trajectory is likely to be far away. The first trajectory is re-colored green (and the second red) according to the probability that the other trajectory is closer than 100 meters (resulting in yellow at locations through which both trajectories move simultaneously).

Solutions for the detection of many movement patterns have been provided in the past, such as the patterns encounter, following, and regular visits. In particular, these have been formally defined and efficient algorithms for detecting them have been given for the linear movement model. Here we formulate them in a meaningful way in a probabilistic setting. For this we provide in Section 2 basic building blocks to deal with problems concerning quantities like position, distance, speed and direction.

We show how these building blocks can be combined to describe movement patterns in Section 3. In particular, we show how to formulate and compute the following patterns under the BBMM: encounter, avoidance and attraction, following, and regular visits. We selected these patterns to demonstrate the use of the various basic building blocks, and because these are commonly needed patterns when analyzing interaction in the setting of movement ecology. We evaluate our framework by the example of encounters between groups of wild primates, specifically vervet monkeys. Computing movement patterns using Brownian bridges is computationally expensive, and we tackle the computational task by parallel computations as described in Section 4. Nowadays, graphics hardware (GPU) offer a widespread, relatively low-cost possibility for parallel computations. We show how to formulate our tasks such that the computations can be handled efficiently in this way. We provide an efficient implementation in C++ and CUDA.

From the application-perspective it is crucial that these methods are also accessible in a widely used analysis framework. In the context of movement ecology \(^1\), an environment for statistical computing and modeling, is commonly used. To this end, we provide a frontend to access our framework from within R. In Section 5 we present a study demonstrating the application of our framework in movement ecology. We use movement data of four neighboring groups of wild vervet monkeys. The data contains one sample every daytime hour and was captured over a period of about three month. In this study we compare the detection of encounters using the linear motion model and the BBMM.

**Related work.** The movement pattern *encounter* was already considered in one of the first papers to study movement patterns in trajectory data [23]. It is basically defined as a set of trajectories that are in the same disk at the same time. For the patterns *attraction* and *avoidance* we are not aware of similar prior work. Avoidance has been considered recently in terms of a swift change of movement to move out of each other’s way [24]. We are here interested in avoidance (and attraction) less as a reactive movement but in terms of avoiding (or seeking) regions in which another entity is. Note that this avoidance (or attraction) does not need to be caused by the presence of the other entity.

There are several definitions of *following* patterns, a direction-based [3] and a path-based definition [6]. In the first definition one entity is following another, if both entities move in a similar direction, and the leading entity is in a cone in front of the following one. In the second definition one entity is following the other if it visits approximately the same locations but with a small, variable time shift. Algorithms for detecting *regular visits* [14] find reoccurring visits of an entity to a certain region. Besides these movement patterns, many more have been considered. A taxonomy of movement patterns is given in [15]. All of the patterns above are also described (in the linear movement model) in a recent survey on algorithmic movement analysis [17].

Movement ecology studies the movement of microorganisms, plants and animals. It aims at understanding the causes, mechanisms, patterns, and consequences of movements [25]. In movement ecology random movement models are commonly used to model movements [19, 12], in particular the BBMM is used to interpolate between sample points [8, 18]. A Brownian bridge between two sample points at times \(t_1\) and \(t_2\) is a Brownian walk over the time interval \([t_1, t_2]\) conditioned under the fact that the moving entity is at the locations of the sample points at the respective times. Typically the locations at \(t_1\) and \(t_2\) are assumed to be drawn from a circular normal distribution due to measurement error. Several variants of the BBMM have been considered [4, 21]. Our framework can also be used with these variants. If the sampling is extremely sparse (e.g., only one sample per day) then interpolating between sample points is no longer reasonable and kernel density estimation is frequently used [2, 31]. This then, however, completely ignores motion between sample points, does not account for temporal autocorrelation, and cannot cope with high data-volumes very well (algorithms used to calculate the kernel’s bandwidth fail to minimise). Kernel density estimation can be combined with methods from time geography to counter these shortcomings [16]. Kernel density estimation and random movement models are also used for visualizing movement data [13, 26]. In this paper we focus on data analysis, but would like to note that such an analysis is most powerful when combined with such visualizations.

Brownian bridges are used for tasks like estimating animal home ranges and migration routes, and for evaluating the influence of resource selection on movement [18]. Methods for home range estimation based on the BBMM are provided in the R packages BBMM\(^2\) and adehabitatHR\(^3\) [9]. Our GPU-based implementation considerably speeds up home range estimation when modern graphics hardware is available. More importantly, our framework opens up the possibility to use movement patterns as the ones discussed above within such an analysis.

\(^1\)www.r-project.org/

\(^2\)http://cran.r-project.org/web/packages/BBMM/

\(^3\)http://cran.r-project.org/web/packages/adehabitatHR/
2. BASIC PROPERTIES

In previous analyses the BBMM has been used to obtain the distribution of the location at a given time, which is then typically integrated over time. In this section we use the distribution of location to derive distributions for other basic properties of moving entities: distance between two entities, velocity, speed and direction\(^4\) of an entity. Note that since we are dealing with distributions, we do not directly obtain speed and direction from velocity.

Deriving these distributions analytically is important for computational efficiency: while we could obtain them by numerical integration based on the location(s), such an integration would increase the computation time by a large factor. We present detailed derivations in this section, also to allow to use them as model for deriving properties (possibly also additional properties) under other random movement models. The later sections of this paper can be read independently of these derivations.

2.1 Position and distance

In the BBMM, the distribution of an entity’s position at any point in time is circular normal [8, 18]. The mean of the distribution is linearly interpolated between the two sample points that are closest in time. Typically, the variance is smallest at the measurement points and largest in between.

Let \(t_i\) be the the time of the \(i\)th measurement and \(z_i\) its recorded position. If \(t_i \leq t \leq t_{i+1}\), then the parameters of the distribution at time \(t\) are calculated as follows:

\[
\bar{\mu}(t) = \alpha z_i + (1-\alpha)z_{i+1} \\
\sigma^2(t) = (t_{i+1} - t_i)\alpha(1-\alpha)\sigma_1^2 + (\alpha^2 + (1-\alpha)^2)\sigma_2^2.
\]

Here, \(\alpha = \frac{t-t_i}{t_{i+1}-t_i}\) moves from 0 to 1 as \(t\) moves from \(t_i\) to \(t_{i+1}\), \(\sigma_1^2\) is the diffusion coefficient, defined by the mobility of the animal, and \(\sigma_2^2\) is the measurement uncertainty.

For the distribution of distance consider two entities \(A\) and \(B\) with mean positions \(\mu_A(t)\) and \(\mu_B(t)\) and variances \(\sigma_A^2(t)\) and \(\sigma_B^2(t)\). Then the difference between the positions of \(A\) and \(B\) at time \(t\) again follows a circular normal distribution with mean \(\mu_A(t) - \mu_B(t)\) and variance \(\sigma_A^2(t) + \sigma_B^2(t)\). The absolute value of this circular normal distribution is distributed according to the Rice distribution with parameters \(||\mu_A(t) - \mu_B(t)||\) and \(\sqrt{\sigma_A^2(t) + \sigma_B^2(t)}\) [20].

2.2 Velocity

Velocity (and therefore also speed and direction) are dependent on the temporal scale at which it is derived [22]. Thus instead of deriving a distribution for one point in time (as for distance), we derive a distribution for the average velocity over a time interval. In an analysis the length of the time interval then should be chosen according to the scale of the pattern considered and to the goal of the analysis.

If we have a Brownian bridge \(B\) over the time interval \([t_{i-1}, t_i]\) with measured positions \(\mu_{t_{i-1}}\) and \(\mu_{t_i}\) for the endpoints and we fix the position at a time \(t \in (t_{i-1}, t_i)\) to \(z(t)\), then it has the property that the position distribution in the time interval \((t_{i-1}, t_i)\) is again described by a Brownian bridge \(B'\). The mean position moves linearly from \(\bar{z}(t)\) to \(\mu_{t_i}\). For the measurement uncertainty of the first endpoint of \(B'\) we use \(\sigma^2(t)\) as computed on \(B\) by the Brownian bridge method.

We use this observation to derive a distribution for the average velocity over a time interval \([t_1, t_2]\). If \(t_1, t_2 \in [t_{i-1}, t_i]\), the positions \(X(t_1), X(t_2)\) of an entity at these times cannot be regarded as independent. Instead, we fix \(X(t_1)\) at a position \(\bar{x}\) and then use the method described above to derive a distribution for \(X(t_2)\). We then integrate \(\bar{x}\) over all positions it can assume (i.e. \(\mathbb{R}^2\)) to obtain the distribution of the velocity over this interval.

Let \(V(t_1, t_2)\) be a random variable describing the average velocity of the entity over the time interval \([t_1, t_2]\), i.e. \(V(t_1, t_2) = \frac{X(t_2) - X(t_1)}{t_2 - t_1}\). Define

\[
W(t_1, t_2) := (t_2 - t_1) \cdot V(t_1, t_2),
\]

so \(W\) is the displacement over the time interval \([t_1, t_2]\). We can now derive the distribution of \(W\):

\[
P[W(t_1, t_2) = \vec{v}] = \mathbb{P}[X(t_2) - X(t_1) = \vec{v}] = \int \mathbb{P}[X(t_2) = \vec{x} + \vec{v}|X(t_1) = \vec{x}] d\vec{x} = \int_{\mathbb{R}^2} \phi(\vec{x}; \mu(t_1), \sigma^2(t_1)) \cdot \phi(\vec{x} + \vec{v}; \mu(t_2), \sigma^2(t_2)) d\vec{x}.
\]

Here, \(\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}\) denotes the probability density function (pdf) of a bivariate circular normal distribution with parameters \(\mu\) and \(\sigma^2\), evaluated at \(x\). \(\mu(t)\) is the mean position at time \(t\) conditioned on \(X(t_1) = \bar{x}\). Let \(\beta := \frac{t_{i+1} - t_i}{t_{i+1} - t_i}\). Then,

\[
\mu(t) = (1 - \beta)\bar{x} + \beta\mu(t_i).
\]

The expression obtained in Equation 1 looks almost like a convolution of two normal distributions, but in order to really be a convolution, the second factor has to be of the form \(f(\vec{c} - \vec{x})\) for some function \(f\) and constant \(\vec{c}\). We can obtain this form using a few variable transformations and moving the (non-constant) mean to the operand part:

\[
\phi(\vec{x} + \vec{v}; \mu(t_2), \sigma^2(t_2)) \cdot \phi(\vec{x}; \mu(t_1), \sigma^2(t_1)) = \phi(\vec{x} + \vec{v} - \beta \bar{x} + \beta \mu(t_i); 0, \sigma^2(t_2)) = \phi(\beta (\vec{v} - \bar{x}) + \beta \mu(t_i); 0, \sigma^2(t_2)) = \phi(\beta \mu(t_i) - \bar{x} - \beta \bar{x}; 0, \sigma^2(t_2)) = \phi(\frac{\beta \mu(t_i) - \bar{x}}{\beta}; -\frac{\bar{x}}{\beta}; 0, \frac{\sigma^2(t_2)}{\beta^2}).
\]

So, we can now rewrite the expression from Equation 1 as a convolution:

\[
\int_{\mathbb{R}^2} \phi(\vec{x}; \mu(t_1), \sigma^2(t_1)) \cdot \phi(\vec{x} + \vec{v}; \mu(t_2), \sigma^2(t_2)) d\vec{x} = \int_{\mathbb{R}^2} \phi(\vec{x}; \mu(t_1), \sigma^2(t_1)) \cdot \phi(\mu(t_i) - \frac{\bar{x}}{\beta}; -\frac{\bar{x}}{\beta}; 0, \frac{\sigma^2(t_2)}{\beta^2}) d\vec{x} = \phi(\mu(t_1), \sigma^2(t_1)) \cdot \phi(0, \frac{\sigma^2(t_2)}{\beta^2}) \cdot \left(\mu(t_i) - \frac{\bar{x}}{\beta}\right).
\]
We then use the property that the convolution of two normal distributions is again a normal distribution:

$$
\phi(\mu(t_1), \sigma^2(t_1)) \ast \phi(0, \sigma^2(t_2)) = \phi(\mu(t_1), \sigma^2(t_1) + \sigma^2(t_2)).
$$

So, the pdf of \( W \) can be expressed as:

$$
f(w) = \phi(\bar{w}; \bar{\mu}(t_2) - \mu(t_1), \beta^2 \sigma^2(t_1) + \sigma^2(t_2)),
$$

that is, the displacement is distributed according to a circular normal distribution with mean \( \bar{\mu}(t_2) - \mu(t_1) \) and variance \( \beta^2 \sigma^2(t_1) + \sigma^2(t_2) \). The velocity distribution is obtained by dividing by \( t_2 - t_1 \).

2.3 Speed and direction

Now that we have established that the velocity follows a bivariate circular normal distribution, we use this to derive expressions for the average speed and the average direction of movement over a time interval \([t_1, t_2]\). As in the case of distance the speed is simply distributed according to a Rice distribution with parameters \(|\mu(t_2) - \mu(t_1)|\) and \(\sqrt{\beta^2 \sigma^2(t_1) + \sigma^2(t_2)}\).

For deriving the distribution of direction, it is convenient to work in polar coordinates. Assume \( V \sim N(\bar{\mu}, \sigma^2) \). Let \( d := |\bar{\mu}|, \theta := \operatorname{atan2}(\bar{\mu}) \) represent \( \bar{\mu} \) in polar coordinates. See Figure 2 for an example of this situation.

The probability density of a direction \( \alpha \) may be considered to be the integral of the distribution of \( V \) over an infinitesimally narrow cone with its apex at the origin. Let \( \bar{u}_\alpha \) be the unit vector at an angle \( \alpha \) from the positive x-axis. This gives the following expression for the probability density function of the direction:

$$
f(\alpha) = \int_0^\infty \phi(\bar{u}_\alpha; \bar{\mu}, \sigma^2) r dr = \int_0^\infty \frac{1}{2\pi \sigma^2} e^{-|\bar{\mu}_\alpha - \bar{\mu}|^2 / 2\sigma^2} r dr.
$$

Now, rotate everything by an angle \(-\alpha\) around the origin, such that \( r\bar{u}_\alpha = r\bar{u}_0 = (r, 0) \) runs over the positive x-axis. The new polar coordinates for \( \bar{\mu} \) are now \((d, \beta) := (d, \theta - \alpha)\), so we get the following expression for \( f(\alpha) \).

\[ f(\alpha) = \int_0^\infty \phi(\bar{u}_\alpha; \bar{\mu}, \sigma^2) r dr = \int_0^\infty \frac{1}{2\pi \sigma^2} e^{-|\bar{\mu}_\alpha - \bar{\mu}|^2 / 2\sigma^2} r dr. \]
entities encounter during a time span $[t_0, t_1]$. One might be tempted to formulate this as the probability that an encounter occurred at a time $t$ or all times $t \in [t_0, t_1]$. Such a formulation has the drawback that these probabilities are not easily computed based on the distributions of the underlying properties of the movement, since the probabilities for different values of $t$ are not independent. We therefore instead consider the expected duration of a pattern.

The probability of encounter at time $t$ is determined by evaluating the cumulative distribution function (CDF) of the Rice distribution at the threshold distance. Unfortunately, this CDF cannot be evaluated directly. The expressions for this function involve either an improper integral or an infinite sum, so we can only hope to approximate the CDF exist. These implementations also give tight bounds on the error in the approximation.

To get the expected duration of encounters with distance threshold $d_0$ over a time interval $[0, T]$ we integrate this probability with respect to time:

$$\mathbb{E}(\text{duration of encounters}) = \int_0^T \mathbb{P}(|X_A(t) - X_B(t)| \leq d_0) \, dt. \tag{3}$$

We compute a numerical approximation of this integral.

We define avoidance (and analogously attraction) as trajectories that have much space use in common, but rarely actually encounter each other. That is, both trajectories visit the same locations but at different times. We detect this by comparing the expected duration of encounters using the trajectory data (as above) and using just the utilization distributions. The latter assumes that both entities move independently, so this comparison is a measure of correlation between their trajectories.

The utilization distribution is the probability distribution of the location of an entity. In the BBMM it can be estimated by integrating the distribution of location over the whole duration $T$ of observation, and dividing this integral by $T$. Given utilization distributions (UDs) $u_A(\vec{x})$ and $u_B(\vec{x})$ of two entities, one can compute the expected duration of encounters disregarding temporal information. This model assumes that the locations of both entities are independent. The probability of an encounter with distance threshold $d_0$ conditioned under the location $\vec{y}$ of the first trajectory is therefore

$$\int_{||\vec{x} - \vec{y}|| \leq d_0} u_B(\vec{x}) \, d\vec{x}.$$

The expected duration of an encounter is now $T$ times the probability of an encounter, that is,

$$\mathbb{E}(\text{duration of encounters based on UD}) = T \cdot \int_{\mathbb{R}^2} u_A(\vec{y}) \int_{||\vec{x} - \vec{y}|| \leq d_0} u_B(\vec{x}) \, d\vec{x} \, d\vec{y}.$$

We speak of avoidance if this expected value is considerably larger (that is, by some factor $f_0 > 1$) than the expected duration computed using Equation 3. We speak of attraction, if it is considerably smaller.

### 3.2 Regular visit

Another important pattern is detecting regular visits [14]. A period $p$ and an offset $o$ define a sequence of time points. Given a trajectory $T$ and an area $A$, one can evaluate at each time point whether $T$ is inside $A$. A regular visit pattern is defined as a subsequence of time points where the density of time points having $T$ inside $A$ is sufficiently high. This problem is equivalent to finding the longest dense substring of a bit string, for which efficient algorithms exist [27, 11, 14].

The algorithm presented by Djordjevic et al. [14] can detect the longest dense substring (with density $c$) in a run-length encoded string $s$ with $k$ runs in $O(k)$ time, regardless of the actual length of the string. The algorithm uses a function $f_1(i)$ that represents the number of $1$’s in the first $i$ bits of $s$. A function $f_2(i) := f_1(i) - ci$ is derived from this function. A substring $s[i_1 \ldots i_2]$ of $s$ is dense if and only if $f_2(i_1) \leq f_2(i_2)$. So, finding the longest dense substring in $s$ is equivalent to finding the $i_1, i_2$ with $f_2(i_1) \leq f_2(i_2)$ such that $i_2 - i_1$ is maximal.

In order to efficiently find these $i_1$ and $i_2$, the algorithm computes two envelopes: The lower left envelope $LLE(i) := \min_{j \in \{i-\ldots, n\}} f_2(j)$ and the upper right envelope $URE(i) := \max_{j \in \{i-\ldots, n\}} f_2(j)$. Then, two pointers $i_1$ and $i_2$ are walked along the LLE and URE, respectively. They both start at position 1 and $i_2$ is then advanced until $URE(i_2) < LLE(i_1)$. Then, $i_1$ is advanced until $URE(i_2) \geq LLE(i_1)$ holds again. This process is repeated until both pointers reach $n$. The largest difference that occurred between $i_2$ and $i_1$ during this process is the longest dense substring of $s$.

This algorithm can be adapted to the BBMM. See Figure 4 for an example of the functions and envelopes used by the adapted algorithm. Once again, we can evaluate the probability that $T$ is inside $A$ at a certain time $t$, denoted $\mathbb{P}(T(t) \in A)$, by integrating the position distribution over $A$.

The original algorithm uses a function $f_1(i)$ which is the number of ones seen in the first $i$ bits of the bit string. Instead of counting the number of ones, we keep track of the expected number of ones in the first $i$ bits, i.e. the expected number of times that $T$ is inside $A$ in the first $i$ time points:

$$f'_1(i) := \sum_{k=0}^i \mathbb{P}(T(o + kp) \in A)$$

Running the original algorithm using this modified function, the algorithm produces the longest substring such that the expected density of this substring meets the threshold. So, the trajectory is inside the desired area at least the desired fraction of the time points on average.

![Figure 4: Illustrations of the functions involved in the regular visit algorithm using the BBMM. (a) shows $f_2'(i)$ and the line $c \cdot i$. In (b) we show $f_2(i) = f_2'(i) - c \cdot i$ with its lower left and upper right envelopes.](image-url)
The original algorithm can process a run-length encoded bit string with \( k \) runs in \( O(k) \) time, regardless of the actual length \( n \) of the string. Unfortunately, it is not realistic for many consecutive time points to have exactly identical probabilities, so run-length encoding will not help us here. Instead, we need to evaluate the probability at each of the \( n \) time points.

When we have these probabilities, the rest of the algorithm only needs a constant number of passes over the data to compute the envelopes and find a solution. Therefore finding the longest expected dense substring can be solved in \( O(n) \) time.

### 3.3 Following

Following patterns have been defined in terms of the direction of movement [3] and in terms of the movement paths [6]. We here formulate both of these patterns in the setting of the BBMM.

Andersson et al. [3] define the front-region front(\( e \)) of an entity \( e \) as the wedge-shaped region with its apex at the entity’s position and centered around its movement direction. Its radius is \( r \) and its apex angle is \( \alpha \). If an entity \( e_i \) is in the front region of another \( e_j \) at some time \( t \), \( e_j \) is said to be in front of \( e_i \). They define that \( e_j \) is following \( e_i \) at time \( t \) if \( e_j \) is in front of \( e_i \) and \( \| d_i - d_j \| \leq \beta \), where \( d_i \) and \( d_j \) are the respective movement directions of \( e_i \) and \( e_j \). We illustrate these definitions in Figure 5.

In the linear model it is straightforward to detect when an entity is following another, since both their position and direction are known exactly. In the Brownian bridge model these are not known exactly. We know the position at time \( t \) has a circular normal distribution. The distribution of direction does not give information about the instantaneous movement direction at time \( t \), but rather the average direction of a time interval. Let this interval have length \( \Delta t \) and be centered around \( t \). The distribution of the direction depends on the position of an entity.

If we fix the position for both entities, we get a distribution for the directions and can therefore compute the probability that \( e_j \) is in front of \( e_i \). Let random variables \( X_i \) and \( X_j \) represent the positions of \( e_i \) and \( e_j \) respectively. Define \( d(\vec{x}) \) as the direction of a vector \( \vec{x} \).

Now,\[
\mathbb{P}[e_j \in \text{front}(e) | X_i = \vec{x}_i \land X_j = \vec{x}_j] = \begin{cases} \mathbb{P}[\|d(V_i) - d(\vec{x}_j - \vec{x}_i)\| \leq \frac{\alpha}{2} | X_i = \vec{x}_i] & \text{if } \| \vec{x}_i - \vec{x}_j \| \leq r \\ 0 & \text{otherwise} \end{cases}
\]

Let \( \delta_i \) denote the probability density function of the direction of entity \( i \) as defined in Equation 2. We then get that

\[
\mathbb{P}[\|d_i - d(\vec{x}_j - \vec{x}_i)\| \leq \frac{\alpha}{2} | X_i = \vec{x}_i] = \int_{d(\vec{x}_j - \vec{x}_i) - \frac{\alpha}{2}}^{d(\vec{x}_j - \vec{x}_i) + \frac{\alpha}{2}} \delta_i(\theta) d\theta.
\]

Next we consider the path-based following pattern [6]. Intuitively this definition states that one entity is following another if it reaches approximately the same location but slightly later. Given two entities we consider their trajectories \( A, B \) as function in time that is \( A(t) \) (resp. \( B(t) \)) is the location of trajectory \( A \) (resp. \( B \)) at time \( t \). Now \( B \) is following \( A \) during the time interval \([s_1, t_1]\) with a time shift in \([t_{\text{min}}, t_{\text{max}}]\) and a distance threshold \( d \) if there is an interval \([s_2, t_2]\) and a continuous, piecewise function \( t: [s_1, t_1] \rightarrow [s_2, t_2] \) such that for all \( t \in [s_1, t_1], t \) the distance \( \text{dist}(A(\tau(t)), B(t)) \) between \( A(\tau(t)) \) and \( B(t) \) at most \( d \), and \( \tau(t) \in [t - \tau_{\text{max}}, t - \tau_{\text{min}}] \).

In the linear motion model this pattern can be computed using a constrained free-space diagram [6]. This diagram is the subspace of the product of the parameter spaces of \( A \) and \( B \) of all pairs \((t, t')\) with \( \text{dist}(A(t'), B(t)) \leq d \) and \( t' \in [t - \tau_{\text{max}}, t - \tau_{\text{min}}] \).

In the BBMM we have to deal with a distribution of distances. For any \((t, t')\) the distribution of the distance between \( A(t') \) and \( B(t) \) is a normal distribution assuming that \( A(t') \) and \( B(t) \) are independent normal distribution. As discussed above, we formulate patterns involving a duration in the BBMM using expected duration. We therefore ask for \( s_\tau \) such that the expected length of time that \( \text{dist}(A(\tau(t)), B(t)) \leq d \) is relatively large.

Instead of a path in the constrained free-space diagram we now have to deal with the line integral of probabilities over the product of parameter spaces. This is closely linked to the so called summed/integral Fréchet distance (or average Fréchet distance if divided by the duration) [5, 7]. While no exact algorithm for computing this distance is known, it can be efficiently approximated by sampling \( A(t) \) and \( B(t) \) and by approximating the line integral by a sum [5].
4. COMPUTATIONAL ASPECTS

Many of the methods described above use in one way or another integrals of probability distributions and other functions. In many cases, these integrals cannot be evaluated analytically and so we have to use numerical integration. This often requires a large amount of resources to evaluate such an integral. Numerical integration involves many invocations of the same function with different arguments, as well as adding up the results of these function invocations. Both tasks can be efficiently implemented on a parallel processor like a GPU.

In the following analysis, we use $N$ to denote the number of measurements in the trajectory being analyzed and by $P$ the number of processors available. When numerically integrating a function over time, $T$ denotes the number of time steps. Similarly, $G$ is the size of the grid when integrating a function over space.

For the basic properties discussed in Section 2, an important operation is finding the parameters for their distribution at a specific time $t$. For example, we might want to obtain the mean and variance for the position of an entity. To obtain these parameters, we must first identify in which Brownian bridge $t$ lies, i.e. find the index $i$ such that $t_i \leq t < t_{i+1}$.

Since a trajectory is ordered on time, we can divide it into $P$ equal parts and use binary search on each part in parallel to obtain $i$ in $\Theta(\log \frac{2}{P})$ time. Then we have to interpolate between measurements $i$ and $i+1$ to obtain the value at time $t$. This can be done in constant time for all properties, so obtaining the parameters of the distribution at time $t$ takes $\Theta(1 + \log \frac{N}{P})$ time.

Another important operation is computing the expected value for some metric, for example the expected duration of encounters described in Section 3.1. We divide the time to be integrated over in $P$ equal chunks.

Every processor has to identify in which bridge to start working, taking $\Theta(\log N)$ time. Then each processor has to evaluate the integrand $\frac{1}{P}$ times and add up the results, which requires $\Theta\left(\frac{N}{P}\right)$ time, assuming the integrand can be evaluated in constant time. Finally the results of all $P$ processors must be added together, which can be done in $\Theta(\log P)$ time. Therefore integrating a function over time takes $\Theta(\log N + \log P + \frac{N}{P})$ time. Similarly we can integrate a function over space in $\Theta(\log N + \log P + \frac{N}{P})$ time.

An operation that is not covered by this generic analysis is determining whether an entity follows another according to the model by Andersson et al., described in Section 3.3. This expression contains four nested integrals. The inner integral does not need to be fully evaluated for every step of the third integral, but instead its previous value can be updated in constant time.

The total number of operations in these integrations is $\Theta(G d^2 A(\alpha + \beta))$, where $A$ is the angular separation of the steps in the inner integrals. This algorithm is perfectly suitable to be run in parallel, so the running time is $\Theta\left(\frac{G d^2 A(\alpha + \beta)}{P}\right)$.

For the path-based following pattern, we refer to the discussion in Section 3.3: The computations for this pattern boil down to computing the summed Fréchet distance, which can be done efficiently when using numerical approximation as discussed in [5].

5. A MOVEMENT ECOLOGY CASE STUDY

The applicability of the framework outlined above is demonstrated by a direct implementation into a pilot study on the movement ecology of wild and free-ranging vervet monkeys (C. pygerythrus).

Vervet monkeys belong to a genus of social primate that is distributed throughout most of sub-Saharan Africa [28]. They live in stable, mixed-sex groups of typically around 25 animals that contain multiple adult males and females along with their offspring. Previous work on home range selection and space use has revealed that vervet monkeys prefer areas of elevated primary productivity [28] and that movement is strongly affected by the distribution of food, surface water and perceived predation risk [30]. Given the current, rapid advances in telemetric technology (with the development of rugged, lightweight location-aware devices), it has recently become feasible to undertake a dedicated study on the simultaneous movement of several neighbouring groups and, more specifically, on how the space use of one group affects (and is itself affected by) the space use of other groups. Viewed in the light of intergroup conflict over limiting resources, this has the potential to generate novel insights into the ecological causes and mechanisms underlying the high degree of within-group cooperation observed in the context of between-group competition, typical of many primate species. Moreover, by comprehending the socio-ecological dynamics of between-group competition over space in non-human primates, we might also further our understanding of the evolutionary roots and ecological causes of primitive warfare (i.e. territorial disputes over resources and space between neighbouring groups) in our own species.

The dataset used here was collected on four habituated groups of vervet monkey on privately owned land in KwaZulu-Natal, South Africa, between 1 st October 2010 and 31 st January 2011 (see Figure 6). Digital telemetry collars, programmed to obtain a GPS-coordinate at hourly intervals throughout the animals’ daily activity period (05:00 – 19:00 o’clock), were deployed on a single adult female within each group. Since groups move as a coherent unit, relocations obtained on this single individual within each group, were taken to be representative of the movement of each of the respective groups. Based on local visibility in the field and

![Figure 6: Spatial distribution of the sample points of the 4 groups](image-url)
the observed behaviour of animals, we defined the maximal distance between groups at which they were considered to be engaged in an encounter ($d_0$ in Section 3.1) to be 100m.

In Figure 7 we show the total duration of encounters between each pair of groups detected by both the linear motion model and the BBMM. Most groups have little to no interaction, but the pair BD-NH stands out here. For certain pairs of groups like NH and RK, the BBMM reports some encounter time while the linear model does not. This is to some extent expected since the BBMM always assigns a positive (although possibly very small) probability of encounter. However, the difference cannot be solely attributed to such small probabilities. The expected location time can rather be mostly attributed to cases where the location measurements of the animals reach a distance that is slightly over 100 meters. The BBMM then assigns a reasonable probability of encounter, while the linear model just states there is no encounter. For instance, Figure 1(b) shows the groups NH and RK on a day on which the distance between their measurements comes close to 100 meters. On that day their expected encounter time in the BBMM is over 9 minutes. The reverse situation also happens: When the linear model detects an encounter, the BBMM still has a positive probability that there is no encounter. This explains why the linear model reports more encounter time than the BBMM for BD-NH.

From Figure 7 it is apparent that there are great differences in the dyadic encounter rates between groups. For example, BD group spent a lot of time in encounters with NH group, but in contrast, never encountered RK group. This striking difference in the duration of encounters between dyads of groups of course partially reflects the fact that the home ranges of some groups overlapped more in space than those of others: Figure 8 shows that groups BD and NH shared large parts of their monthly home ranges (demarcated by 99% volume isopleths calculated from the summed daily BBMMs over, in this case, December 2010), whereas BD and RK did not exhibit any such spatial overlap at all. However, this difference in proportional home range overlap can only partly account for the observed differences in dyadic encounter rates and durations. Other factors, both social and ecological, need to be considered as the overlap between BD and AK was in the same order of magnitude as the overlap between BD and NH, yet the total encounter duration was many times larger for the latter dyad.

In this respect, it is worthy to note that the observation period (October-January) coincides with the onset of the rainy season in Kwazulu-Natal. This implied that, after half a year of relative food scarcity, highly preferred food items started sprouting after the arrival of the first rains and the area that was among the first to benefit from this, was the tall riverine vegetation situated in a loop of the river that falls exactly within the BD and NH overlap area. Although no quantitative data on food availability are available over the period of this pilot study, field observations indeed strongly suggest that BD and NH were competing fiercely over this new food source.

![Figure 8: Estimation of monthly home ranges of the 4 groups computed using the Brownian bridge motion model.](image)

![Figure 9: Days with large expected encounter time.](image)

![Figure 10: Daily trajectories for two groups with encounters (BD-NH day 1417).](image)
Following the consideration of total duration of encounters between the different group dyads, we now turn our attention to days on which long encounters are expected to have occurred. Figure 9 shows pairs of daily trajectories for which either the linear model reports an encounter, or the BBMM indicates an expected duration of encounter exceeding 15 minutes. For each of these days we present the identities of the groups concerned and show bar charts of the expected duration of encounter according to both movement models.

Focusing on a particular encounter, BD-NH day 1477 from Figure 9 is displayed in geographic space in Figure 10. Trajectories of both groups are given for both the linear movement model and the BBMM and are superimposed upon a high resolution satellite image, indicative of local primary productivity (darker shades of grey represent more photosynthetically active vegetation, with the river standing out as a light white band). This superimposition upon ecological variables of potential interest may prove highly informative in establishing which environmental conditions are associated with frequent or long encounters between groups.

Figure 11 shows the distance between the groups over time and Figure 12 the probability of encounter over time. Starting at approximately 15:00 hours until the end of the measurement period, the groups moved along approximately parallel trajectories and both models recognize this as a likely encounter. Around 18:00 the measured distance is slightly larger than 100 meters and thus the linear model states that no encounter occurs at that time. In reality it is much more likely that the groups were still moving together, but for some reason at the measurement time the distance was somewhat larger. The BBMM still reports a reasonable probability that an encounter is in progress at that time.

In Figure 11 we see that the distance estimate has a smaller error estimate at the measurement times (every full hour) than in between these times, as is expected with the BBMM. We also note that the distance estimate in the BBMM goes up when the mean is small compared to the standard deviation. This implies that the probability of encounter decreases between measurements, when the variance is larger. This effect can be clearly seen in Figure 12.

**Implementation details.** We implemented the algorithms for detecting encounters in R, with the crucial parts implemented in C++ for performance reasons. In particular we extended the 1traj class in the package adehabitatLT [10] so it stores the measurement uncertainty and diffusion coefficient for each measurement. This allows to use dynamic bridges with different values for different measurements. We are using a measurement error that was recorded by the GPS collars. The diffusion coefficient was fixed for each group and estimated using the maximum likelihood method [18].

We then implemented various functions for analyzing these trajectories. These include functions to compute the total duration of encounters and the distance distribution at any point in time. Given the distance distribution, we can also compute the probability that an encounter occurs.

It is interesting to note that the functions that compute the encounter duration for both models have nearly identical running times. This is somewhat surprising, since the linear model only has to solve one quadratic equation for each measurement. The BBMM performs numerical integration with multiple steps for each bridge. In every step the cumulative density function of the Rice distribution is evaluated, which is costly. When computing results like the utilization distributions shown in Figure 10(b), BBMM computation takes significantly longer. In that case it has to repeat a similar operation for each grid cell. The integration of our framework in R makes the generation of plots as in Figures 7, 9, 11 and 12 a simple task using the functionality given in R.

**6. CONCLUSION**

In this paper we developed a framework for algorithmic movement analysis using the Brownian bridge movement model (BBMM). Movement patterns composed from basic movement properties can be computed efficiently in this framework. Some movement patterns that we considered cannot be composed in such a way, for instance, the regular visits and path-based following. For these patterns the formulation in the BBMM resulted in interesting algorithmic questions of their own right. We expect more interesting research questions to arise when further movement patterns are (re-)formulated in our framework.

Our study on groups of wild and free-ranging vervet monkeys shows that the BBMM can detect interesting patterns that are not detected by the linear model. Instead of relying on the assumption that movement is linear between samples and the location known at all times, the BBMM can detect situations where the condition is likely to be satisfied. For the simultaneous movement data of the case study, the encounter pattern in combination with the BBMM proved a powerful tool for studying movement interaction.
References


