

Stochastic differential equation based on a Gaussian potential field to model fishing vessels trajectories

Pierre Gloaguen* Marie-Pierre Etienne† Sylvain Le Corff§

Abstract

In this paper, a new parametric model continuous in time and space is introduced to analyze trajectory data in ecology. This model assumes that the trajectory of an individual is a solution to a stochastic differential equation. The drift of this equation is defined as the gradient of a potential map which is a mixture of an unknown number of Gaussian shaped functions. Each component of this mixture may be understood as an attractive field characterizing the propensity of the individual to move to certain regions. The parameters of this model are estimated using an Monte Carlo Expectation Maximization algorithm based on the exact algorithm proposed by [BPR06] to sample trajectories exactly distributed as the solution to the stochastic differential equation. The main advantage of this discretization free method is to be efficient even when the data are obtained at irregular times and at a slow rate. The performance of the proposed model and estimation procedure is illustrated using simulated data and true GPS positions of French vessels moving in the English Channel.

1 Introduction

In ecology, studies on the movement of animals provide many insights on the ecological features that explain population-level dynamics. These analyses are crucial to wildlife managers to understand complex animal behaviors [CG06]. In fisheries science, understanding the underlying patterns responsible for spatial use of the ocean is a key aspect of a sustainable management [Cha11].

Both fields promote now large programs to deploy Global Positionning System (GPS) device. For instance we might mention, among others, the Tagging of Pelagic Predators program (TOPP)¹ or the TORSOOI² program for marine animals and the WOLF GPS³ or Elephant without Borders⁴ programs for terrestrial wildlife. In the European Union, since the 1st of January 2012, the fishing vessels above 12m are mandatory equipped with a Vessel Monitoring System which has become a standard tool of fisheries monitoring and control worldwide. As a result such programs produce large amount of trajectory data. These data sets have been largely used to understand, explain and potentially predict animals/vessels movements.

Those tasks require modeling and analysis of the GPS tracks data but may have different objectives. A first objective is to segment the whole trajectory in homogeneous regions which are to be linked with behavioural activities. This is classically addressed using Hidden Markov Models [JBT13, GMR⁺15] or change point detection approaches [BB08]. A second objective is the construction/definition of a land/space use map, defined as Utilization Distribution (UD) of the individual, using GPS data. Two main approaches are used to build these UD maps. First, Nonparametric kernel methods usually require assumptions such as independent data that depend strongly on the sampling scheme of the data (although these assumptions have been relaxed to account for autocorrelation in some recent works [CFM⁺15]). Another approach, based on Brownian Bridge techniques, assumes that the trajectory is a Brownian motion and then builds the UD by integrating the probability of presence over time.

These two approaches rely on unrealistic assumptions regarding the movement dynamics. In this paper, it is assumed that observed trajectories of an individual are direct consequences of the environment attractiveness, hereafter called potential. We develop a method based on continuous

*IFREMER, Ecologie et Modèles pour l’Halieutique, BP 21105, 44311 Nantes Cedex 03, France.

†AgroParistech, UMR MIA 518, F-75231 Paris, France.

‡INRA, UMR MIA 518, F-75231 Paris, France.

§Laboratoire de Mathématiques d’Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France.

¹<http://www.gtopp.org/>

²<http://www.torsooi.com/>

³<http://www.wolfgps.com/>

⁴<http://www.elephantwithoutborders.org/tracking.php>

time and continuous space stochastic modeling to estimate potential maps. The random process of interest $(X_t)_{t \geq 0}$ describes the position of an individual continuously in time and space but discretely observed. This process is assumed to be a solution to a stochastic differential equation (SDE) which depends on the environment potential. This statistical framework allows to (a) estimate the parameters of the SDE and (b) sample trajectories of individuals to predict future behaviors.

The use of a special form of SDE to model animal movement have been studied in [JLLD08] and [HB13]. These papers focus on the mean-reverting Ornstein-Uhlenbeck process and its extensions. The Ornstein Ulhenbeck process has a known distribution that facilitates parameters estimation. However, it is a quite restrictive class of model, and its extensions make the estimation framework more complex. [PAJK04] and [PAW13] consider the special case where the drift of the SDE is the gradient of a potential function which is a weighted sum of different sources that define attractive (resp. repulsive) regions where the individuals are likely (resp. unlikely) to move. As the solution to the corresponding SDE does not have an explicit density, the drift function is estimated by applying an Euler scheme to interpolate trajectories between two consecutive observations. This introduces an intrinsic bias in the estimation procedure, and might not be well suited for tracking data that can be sparsely sampled.

In this paper, a SDE based on a potential function is proposed to model the position of an individual at each time step t . The drift of the SDE is a mixture of Gaussian shaped functions, with unknown weights, centers and shapes, which represent the attractive regions where the species/fishing vessels are likely to travel.

The aim of this paper is to find the maximum likelihood estimator (MLE) for these attractive regions using GPS data. Although the SDE has no explicit solution, considering G independent animal/vessel trips $\mathbf{X}^1, \dots, \mathbf{X}^G$, a trip \mathbf{X}^g being a sequence of observations $\mathbf{X}^g = (X_0^g, \dots, X_{n_g}^g)$ sampled at times $t_0^g, \dots, t_{n_g}^g$, the parameters of the SDE (weights, centers and shapes of the attractive regions) can be estimated using an Expectation Maximization (EM) based algorithm. The E step is approximated by Monte Carlo methods, using the exact algorithms introduced in [BPR06] and [BR05] for an exact sampling of the SDE. The M step is performed using the gradient free CMAES approach described in [HO01]. As this proposed method does not rely on a discrete scheme to approximate the true process, the estimation error of the MLE only depends on the quality of the Monte Carlo approximation of the E step.

The paper is organized as follows. In Section 2, the target SDE is introduced and the EM procedure based on independent trips to estimate the parameters is displayed in Section 3. Performance of the proposed algorithm is assessed in Section 4.1 with simulated data and in Section 4.2 using a real data set. Technical results to apply the EA algorithm of [BPR06, BR05] are postponed to Appendix A and B.

2 Model and objectives

The goal of this paper is to propose a model which allows to identify regions of high attractiveness for an individual using GPS tagging. Those regions may then be ecologically interpreted and understood as feeding zones for different animals or high concentration of commercially interesting fishes for fishing vessels. The movement is modeled using a SDE on \mathbb{R}^2 . The drift is the gradient of a potential map which value at location x represents the attractiveness. The diffusion term is assumed to be a constant scalar matrix. This apparently quite restrictive assumption is motivated by technical reasons but seems also quite reasonable due to a lack of biological information on the stochastic part of the movement.

Formally, the model is defined as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let \mathcal{C} be the set of continuous maps from $[0, T]$ ($T > 0$ is a fixed time horizon) to \mathbb{R}^2 , called the Wiener space, endowed with the usual σ -algebra \mathcal{C} generated by cylinders. In the following, we consider the unique measure \mathbb{W}_T on the Wiener space such that the coordinate process $(W_t)_{0 \leq t \leq T}$ is a standard Brownian motion and the natural filtration $\{\mathcal{F}_t\}_{t \geq 0}$. Each observed trajectory \mathbf{X}^g is a realization of the position process $(X_t)_{t \geq 0}$ which is assumed to be a solution to the following homogeneous SDE:

$$X_0 = X_0^g \quad \text{and} \quad dX_t = b_\theta(X_t)dt + \gamma dW_t, \quad (1)$$

where $\gamma \in \mathbb{R}$ is the diffusion coefficient and the drift function b_θ is defined as follows : $b_\theta := \nabla P_\theta$,

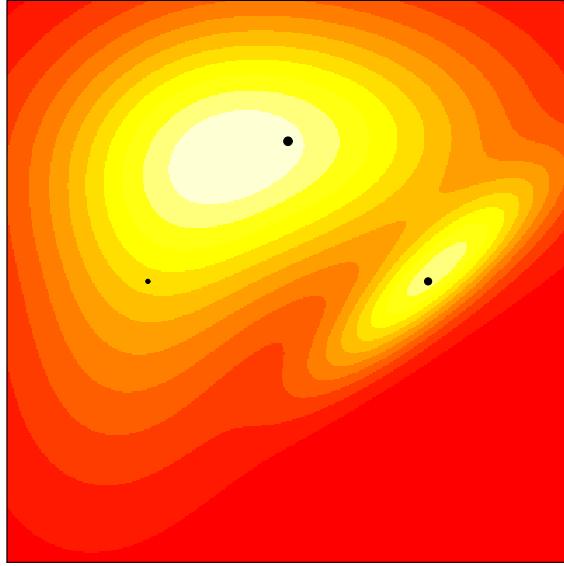


Figure 1: One example of potential map (when $K = 3$). Each black dot represents the position of μ_k ($k = 1, 2, 3$), the dot size being relative to the weight π_k .

where $P_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given, for all $x \in \mathbb{R}^2$, by

$$\begin{aligned} P_\theta(x) &:= \sum_{k=1}^K \pi_k \varphi_k^\theta(x), \\ \varphi_k^\theta(x) &:= \exp\left(-\frac{1}{2}(x - \mu_k)^T C_k (x - \mu)\right). \end{aligned} \quad (2)$$

where $\pi_k \in \mathbb{R}$, $\mu_k \in \mathbb{R}^2$, and C_k is a positive-definite matrix. The parameter θ contains all the unknown quantities we wish to estimate: $(\mu_k)_{1 \leq k \leq K}$, $(C_k)_{1 \leq k \leq K}$ and $(\pi_k)_{1 \leq k \leq K}$ and we will discuss either γ has to be estimated or not. μ_k , C_k and π_k are respectively location, shape and weight parameters of the k -th attractive zone. The number of attractive zones K is first assumed to be known. Therefore, for all $x \in \mathbb{R}^2$,

$$b_\theta(x) = - \sum_{k=1}^K \pi_k \varphi_k^\theta(x) C_k (x - \mu_k). \quad (3)$$

Since b_θ is Lipschitz and γ is constant, the solution to this SDE exists and is unique.

Figure 1 shows one possible example of map of space use, the gradient of which defining the drift b of the proposed SDE. The model therefore captures the idea of attractive zones we aim at identifying. The parametric form of (2) provides a smooth and flexible framework to describe different potential maps. The potential map is chosen positive (but could be chosen negative), but has no constraint to integrate to 1. For these reasons, the map is not a probability distribution. However, as the potential is supposed to be bounded, it might be understood as a measure of attractiveness.

The assumption on the diffusion parameter is quite strong. In a more general context we could have defined the diffusion term as a matrix Γ but we would need to ensure that the proposed model is such that the function $x \mapsto \Gamma^{-1}b(\Gamma x)$ is conservative, i.e. such that there exists

$$H_\theta : \mathbb{R}^2 \rightarrow \mathbb{R} \text{ satisfying } \nabla H_\theta = \Gamma^{-1}b_\theta(\Gamma x). \quad (4)$$

In the model described by (1) and (3), the mapping $x \mapsto \gamma^{-1}b(\gamma x)$ is conservative with

$$H_\theta : x \mapsto \sum_{k=1}^K \pi_k \varphi_k^\theta(\gamma x) / \gamma. \quad (5)$$

This conservative property is crucial to apply the exact algorithm 1 of [BPR06], which allows to sample skeleton of trajectories exactly distributed as the solution to (1). This assumption could

be modified in the case of a non constant diffusion matrix $x \mapsto \Gamma(x)$ at the cost of a lot of technical complications.

This model depends on the parameters γ , $(\pi_k)_{1 \leq k \leq K}$, $(\mu_k)_{1 \leq k \leq K}$ and $(C_k)_{1 \leq k \leq K}$. We will discuss two cases whether γ is known or not in Section 3. As the C_k 's are symmetric positive definite matrices, $6K$ or $6K + 1$ parameters have to be estimated. The aim of this paper is then to estimate these parameters using a set of G independent and partially observed realizations $(\mathbf{X}^g, g = 1, \dots, G)$ of the process (X_t) . The $n_g + 1$ observations of the g -th realization are observed at times $(t_0^g, \dots, t_{n_g}^g)$ and are denoted $(X_0^g, \dots, X_{n_g}^g)$.

Parameter estimation for diffusion processes is a complicated task due to the unavailability of the transition density of the Markov process defined by (1), see [Sor04] and the references therein for a recent survey. Many methodologies introduce an approximation of this transition density, mostly based on Euler scheme, see for instance [AS08] for an explicit approximation which converges uniformly on the parameter space to the true transition density. See also [DG02, Ped95] for methods relying on approximate Monte Carlo maximum likelihood or [ECS01, Era01, RS01] for Markov chain Monte Carlo methods combined with data augmentation. In this paper, the estimation procedure is based on the exact algorithm introduced in [BPR06]. This algorithm allows to sample skeletons of trajectories exactly distributed as the finite dimensional distribution of the target SDE (1). These skeletons can then be used to obtain unbiased Monte Carlo estimates of the intermediate quantity of the Expectation Maximization algorithm to define maximum likelihood estimates of θ , as described in [BPRF06].

3 Expectation Maximization based procedure

This section provides an algorithm to estimate the parameter θ using a set of G independent trips $\mathbf{X}^1, \dots, \mathbf{X}^G$. Statistical inference based on a set of observations usually requires the finite-dimensional distributions of the process (X_t) to be available. However, in the context of this paper, thanks to the results of [BR05, BPRF06], the likelihood of the complete path (X_t) with respect to a reference measure on $(\mathcal{C}, \mathcal{C})$ can be obtained easily using Girsanov theorem even if the finite dimensional distribution of $(X_{t_0}, \dots, X_{t_n})$ is not explicitly available. This setting is conducive to the use of the EM algorithm proposed by [DLR77] which allows to perform maximum likelihood estimation when the joint distribution of the observations and some additional missing data is available. Starting with an initial estimate θ_0 , the EM algorithm produces iteratively a sequence $(\theta_p)_{p \geq 0}$ that converges toward a local maximum of the likelihood of the observations. The EM based algorithm proposed in this paper relies on the Monte Carlo EM procedure presented in [BPRF06].

E-step

For the sake of clarity, the E-step is first presented assuming that the diffusion γ is known, the general case is developed in the following section.

Case where the diffusion coefficient γ is known: in this case, the parameter to be estimated is

$$\theta := \{(\pi_k)_{1 \leq k \leq K}; (\mu_k)_{1 \leq k \leq K}; (C_k)_{1 \leq k \leq K}\}.$$

For all θ , the law of the process solution to (1) is absolutely continuous with respect to the law of the driftless diffusion γdW_t . In this case, the dominating measure on $(\mathcal{C}, \mathcal{C})$ does not depend on θ as γ is known so that maximum likelihood estimation of θ is possible. The procedure proposed in [BPRF06] is based on rejection sampling and uses a reparametrization of the observations to simplify the dominating measure: the process $(X_t)_{0 \leq t \leq T}$ is transformed into a new diffusion process $(Y_t)_{0 \leq t \leq T}$ with unit diffusion coefficient. Define the Lamperti transform $\eta : x \mapsto \gamma^{-1}x$. By Ito's formula, the process $(Y_t) = \eta(X_t)_{0 \leq t \leq T}$ satisfies $Y_0 = \gamma^{-1}X_0$, $Y_T = \gamma^{-1}X_T$ and

$$dY_t = \alpha_\theta(Y_t)dt + dW_t, \quad (6)$$

where α_θ is given by

$$\alpha_\theta : \mathbb{R}^2 \mapsto \mathbb{R}^2 \quad (7)$$

$$x \mapsto \gamma^{-1}b_\theta(\gamma x) = -\sum_{k=1}^K \pi_k \varphi_k^\theta(\gamma x) \gamma^{-1} C_k^{-1}(\gamma x - \mu_k).$$

Let \mathbb{Q}_T^θ be the law of $(Y_t)_{0 \leq t \leq T}$ on $(\mathcal{C}, \mathcal{C})$ when the SDE is parameterized by θ . By Girsanov formula, \mathbb{Q}_T^θ is absolutely continuous with respect to the Wiener measure \mathbb{W}_T on $(\mathcal{C}, \mathcal{C})$ and its Radon-Nikodym derivative is given by:

$$\frac{d\mathbb{Q}_T^\theta}{d\mathbb{W}_T}(\omega) = \exp \left\{ \int_0^T \alpha_\theta(\omega_s) d\omega_s - \frac{1}{2} \int_0^T \|\alpha_\theta(\omega_s)\|^2 ds \right\}.$$

Then, applying Ito's formula,

$$\frac{d\mathbb{Q}_T^\theta}{d\mathbb{W}_T}(\omega) = \exp \left\{ H_\theta(\omega_T) - H_\theta(\omega_0) - \frac{1}{2} \int_0^T [\Delta H_\theta(\omega_s) + \|\alpha_\theta(\omega_s)\|^2] ds \right\},$$

where H_θ is defined by (5) and

$$\Delta H_\theta : x \mapsto \frac{\partial \alpha_{\theta,1}}{\partial x_1}(x) + \frac{\partial \alpha_{\theta,2}}{\partial x_2}(x). \quad (8)$$

Therefore, the log-likelihood function of a complete path is given by:

$$L(\omega, \theta) := H_\theta(\omega_T) - H_\theta(\omega_0) - \frac{1}{2} \int_0^T [\Delta H_\theta(\omega_s) + \|\alpha_\theta(\omega_s)\|^2] ds.$$

For each trip $1 \leq g \leq G$, the observations \mathbf{X}^g are transformed into \mathbf{Y}^g where for all $0 \leq k \leq n^g$, $Y_k^g = \gamma^{-1} X_k^g$. Given a current estimate θ_p , the E-step consists in the computation of the intermediate quantity $\theta \mapsto Q(\theta, \theta_p)$ given, as the transformed trips $(\mathbf{Y}^g, g = 1, \dots, G)$ are independent, by

$$\begin{aligned} Q(\theta, \theta_p) &:= \sum_{g=1}^G \mathbb{E}_{\theta_p} [L(Y, \theta) | \mathbf{Y}^g], \\ &= \sum_{g=1}^G \left\{ H_\theta(Y_{n^g}^g) - H_\theta(Y_0^g) - \frac{1}{2} \mathbb{E}_{\theta_p} \left[\int_0^{t_{n^g}^g} [\Delta H_\theta(Y_s) + \|\alpha_\theta(Y_s)\|^2] ds \middle| \mathbf{Y}^g \right] \right\}, \\ &= \sum_{g=1}^G \left\{ H_\theta(Y_{n^g}^g) - H_\theta(Y_0^g) - \frac{1}{2} \sum_{j=1}^{n^g} \mathbb{E}_{\theta_p} \left[\int_{t_{j-1}^g}^{t_j^g} [\Delta H_\theta(Y_s) + \|\alpha_\theta(Y_s)\|^2] ds \middle| \mathbf{Y}^g \right] \right\}, \end{aligned}$$

where $\mathbb{E}_{\theta_p}[\cdot | \mathbf{Y}^g]$ denotes the conditional expectation under the law of the process (Y_t) on $(\mathcal{C}, \mathcal{C})$ given \mathbf{Y}^g when the SDE is parameterized by θ_p .

The conditional expectations required to compute the intermediate quantity $\theta \mapsto Q(\theta, \theta_p)$ are not available analytically but, as noted by [BPRF06],

$$\int_{t_{j-1}^g}^{t_j^g} [\Delta H_\theta(Y_s) + \|\alpha_\theta(Y_s)\|^2] ds = (t_j^g - t_{j-1}^g) \mathbb{E}_{\theta_p} [\Delta H_\theta(Y_{U^{g,j}}) + \|\alpha_\theta(Y_{U^{g,j}})\|^2 | (Y_t)_{0 \leq t \leq T}],$$

where $U^{g,j}$ is independent of $(Y_t)_{0 \leq t \leq T}$ and uniformly distributed on $[t_{j-1}^g, t_j^g]$. Then, the quantity $Q(\theta, \theta_p)$ may be estimated by Monte Carlo simulations. For all $1 \leq g \leq G$, $1 \leq j \leq n^g$ and all $1 \leq i \leq N_j^g$,

- (i) simulate $(U_k^{g,j,i})_{1 \leq k \leq M_j}$ independently and uniformly on $[t_{j-1}^g, t_j^g]$;
- (ii) conditional on Y_{j-1}^g and Y_j^g , sample a skeleton $Y^{g,j,i}$ at time instances $(U_k^{g,j,i})_{1 \leq k \leq M_j^g}$.

Then, $Q(\theta, \theta_p)$ is estimated by $Q^N(\theta, \theta_p)$ where

$$Q^N(\theta, \theta_p) := \sum_{g=1}^G \left[H_\theta(Y_{n^g}^g) - H_\theta(Y_0^g) - \frac{1}{2} \sum_{j=1}^{n^g} \frac{t_j^g - t_{j-1}^g}{M_j^g N_j^g} \sum_{i=1}^{N_j^g} \sum_{k=1}^{M_j} \left\{ \Delta H_\theta \left(Y_{U_k^{g,j,i}}^{g,j,i} \right) + \left\| \alpha_\theta \left(Y_{U_k^{g,j,i}}^{g,j,i} \right) \right\|^2 \right\} \right].$$

The procedure to sample each $Y_{U_k^{g,j,i}}^{g,j,i}$ given Y_{j-1}^g and Y_j^g is the Exact Algorithm 1 (EA1) of [BPR06]. It is detailed in Appendix A for completeness along with technical results in Appendix B for the specific implementation details to be applied to the model presented in this paper.

Case where the diffusion coefficient γ is unknown: in this case, the parameter to be estimated is

$$\theta := \{(\pi_k)_{1 \leq k \leq K}; (\mu_k)_{1 \leq k \leq K}; (C_k)_{1 \leq k \leq K}; \gamma\}.$$

Then, the transformation η to obtain the unitary diffusion (6) depends on θ . For all $1 \leq g \leq G$, the set $(Y_0^g(\theta), \dots, Y_{n^g}^g(\theta))$ is not directly observed but is a function of the unknown parameter θ . [BPRF06] suggested to use a second path transformation to define a dominating measure which does not depend on θ . Define, for all $1 \leq g \leq G$, $1 \leq j \leq n^g$ and all $s \in [t_{j-1}^g, t_j^g]$,

$$\dot{Y}_s^g(\theta) := Y_s^g(\theta) - \left(1 - \frac{s - t_{j-1}^g}{t_j^g - t_{j-1}^g}\right) Y_{j-1}^g(\theta) - \frac{s - t_j^g}{t_j^g - t_{j-1}^g} Y_j^g(\theta). \quad (9)$$

The transformation (9) maps $(Y^g(\theta))_{t_{j-1}^g \leq s \leq t_j^g}$ onto the diffusion bridge $(\dot{Y}^g)_{t_{j-1}^g \leq s \leq t_j^g}$ starting and ending at 0 for all θ . The law of this transformed process $(\dot{Y}^g)_{t_{j-1}^g \leq s \leq t_j^g}$ is absolutely continuous with respect to the law of the Brownian bridge on $[t_{j-1}^g, t_j^g]$ starting and ending at 0. The inverse transform of (9) is given by:

$$f_\theta(\dot{Y}_s^g(\theta')) := \dot{Y}_s^g(\theta') + \left(1 - \frac{s - t_{j-1}^g}{t_j^g - t_{j-1}^g}\right) Y_{t_{j-1}^g}^g(\theta) + \frac{s - t_j^g}{t_j^g - t_{j-1}^g} Y_{t_j^g}^g(\theta). \quad (10)$$

The complete path used in the EM algorithm is now $\{\dot{Y}_t^g\}_{0 \leq t \leq T}; \mathbf{Y}^g\}$ and the intermediate quantity of the EM algorithm is, by [BPRF06, Lemma 2],

$$Q(\theta, \theta_p) = \sum_{g=1}^G \left\{ -2n^g \log \gamma + H_\theta(Y_{n^g}^g(\theta)) - H_\theta(Y_0^g(\theta)) + \sum_{j=1}^{n^g} \log \phi_{t_j^g - t_{j-1}^g}(Y_j^g(\theta) - Y_{j-1}^g(\theta)) \right. \\ \left. - \frac{1}{2} \sum_{j=1}^{n^g} \mathbb{E}_{\theta_p} \left[\int_{t_{j-1}^g}^{t_j^g} [\Delta H_\theta(f_\theta(\dot{Y}_s^g(\theta_p))) + \|\alpha_\theta(f_\theta(\dot{Y}_s^g(\theta_p)))\|^2] ds \middle| \mathbf{Y}^g \right] \right\},$$

where ϕ_u is the probability density function of a 2 dimensional $\mathcal{N}(0, uI)$ random variable. Then, following the same steps as in the case where γ is known, the quantity $Q(\theta, \theta_p)$ may be estimated by Monte Carlo simulations. For all $1 \leq g \leq G$, $1 \leq j \leq n^g$ and all $1 \leq i \leq N_j^g$,

- (i) simulate $(U_k^{g,j,i})_{1 \leq k \leq M_j^g}$ independently and uniformly on $[t_{j-1}^g, t_j^g]$;
- (ii) conditional on $Y_{j-1}^g(\theta_p)$ and $Y_j^g(\theta_p)$, draw a skeleton $Y^{g,j,i}$ at times $(U_k^{g,j,i})_{1 \leq k \leq M_j^g}$;
- (iii) compute $\dot{Y}^{g,j,i}(\theta_p)$ at time instances $(U_k^{g,j,i})_{1 \leq k \leq M_j^g}$ by evaluating (9) at $Y_{U_k^{g,j,i}}^{g,j,i}(\theta_p)$.

Then, $Q(\theta, \theta_p)$ is estimated by $Q^N(\theta, \theta_p)$ where

$$Q^N(\theta, \theta_p) := \sum_{g=1}^G \left\{ -2n^g \log \gamma + \sum_{j=1}^{n^g} \log \phi_{t_j^g - t_{j-1}^g}(Y_j^g(\theta) - Y_{j-1}^g(\theta)) + H_\theta(Y_{n^g}^g) - H_\theta(Y_0^g) \right. \\ \left. - \frac{1}{2} \sum_{j=1}^{n^g} \frac{t_j^g - t_{j-1}^g}{M_j^g N_j^g} \sum_{i=1}^{N_j^g} \sum_{k=1}^{M_j^g} \left\{ \Delta H_\theta(f_\theta(\dot{Y}_{U_k^{g,j,i}}^{g,j,i})) + \|\alpha_\theta(f_\theta(\dot{Y}_{U_k^{g,j,i}}^{g,j,i}))\|^2 \right\} \right\}. \quad (11)$$

M-step

In both cases (γ known or unknown), as the function $\theta \mapsto Q^N(\theta, \theta_p)$ cannot be maximized analytically, the M-step is performed numerically. This step is based on the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) introduced in [HO01] which is a derivative-free optimization procedure. The CMA-ES is known to perform well for complex multimodal optimization problems, see e.g. [HK04]. In our case, the constrained version of CMAES should be used but the following parametrization circumvents the constraints problem and the classical version of CMA-ES algorithm is finally used:

- each C_k is a positive definite matrix and may be written

$$C_k = \begin{pmatrix} \exp(2a_1^k) & a_3^k \exp(a_1^k) \\ a_3^k \exp(a_1^k) & \exp(2a_2^k) + (a_3^k)^2 \end{pmatrix},$$

where $a_1^k, a_2^k, a_3^k \in \mathbb{R}$;

- for all $1 \leq k \leq K$, $\pi_k > 0$ and may be written $\exp(\tilde{\pi}_k)$, with $\tilde{\pi}_k \in \mathbb{R}$;
- $\gamma > 0$ and may be written $\exp(\tilde{\gamma})$, with $\tilde{\gamma} \in \mathbb{R}$.

Parameters of the CMA-ES algorithm are tuned according to the heuristics given in [HK04], except for the initial standard deviation at each MCEM step. It is chosen to be increasing from a small to a large value during the first iterations, and then decreasing from this large value to a smaller one for last iterations. This method allows a good exploration of the parameter space without using time consuming adaptative techniques.

Model selection

In the previous section, the number of mixture components K has been assumed to be known and fixed. In the context of mixture models, following [BCG00] the number of components is classically selected according to Integrated Completed Likelihood (ICL) criterion. Even if the shape of the target potential function reminds this context, the missing data in the presented work are the full trajectories and not the component identifier, therefore the ICL criterion can not be easily derived in our context. The number of components is obtained by approximating the Akaike Information criterion by:

$$AIC(\hat{\theta}) = -2 \sum_{g=1}^G L(\mathbf{Y}^g, \hat{\theta}) + 2\dim(\hat{\theta}),$$

where $L(\mathbf{Y}^g, \hat{\theta})$ is the loglikelihood of the g -th observed path. The estimation approach adopted in this paper provides, for a given estimator $\hat{\theta}$, a surrogate for the loglikelihood of the observed paths by computing $Q(\hat{\theta}, \hat{\theta})$ using equation (11). However, we cannot compute $\sum_{g=1}^G L(\mathbf{Y}^g, \hat{\theta})$ as

$$\sum_{g=1}^G L(\mathbf{Y}^g, \hat{\theta}) = Q(\hat{\theta}, \hat{\theta}) - \sum_{g=1}^G \mathbb{E}_{\hat{\theta}} [\log p_{\hat{\theta}}(\mathbf{Y}^{-g} | \mathbf{Y}^g) | \mathbf{Y}^g],$$

where $p_{\hat{\theta}}(\mathbf{Y}^{-g} | \mathbf{Y}^g)$ is the conditional distribution of the missing path given \mathbf{Y}^g when the parameter is $\hat{\theta}$ which is not available analytically. Nevertheless, following [BPR09, Theorem 1], the EA1 algorithm allows to use Monte Carlo simulations to obtain an estimator $L^N(\mathbf{Y}^g, \hat{\theta})$ of the loglikelihood $L(\mathbf{Y}^g, \hat{\theta})$ for all $1 \leq g \leq G$ which in turn defines an estimate of $AIC(\hat{\theta})$ to choose the best model. This procedure is detailed in Appendix C.

4 Experimental results

4.1 Simulated data set

This section illustrates the performance of our procedure using simulated data in the case where the diffusion coefficient γ is unknown. For a given set of parameters, a toy set of trips $\mathbf{X}^1, \dots, \mathbf{X}^G$ satisfying (1) is simulated using the exact algorithm EA1 of [BPR06]. From this data set, the MLE is estimated using the algorithm described in section 3. The estimation is performed with $K = 1$ (7 parameters to estimate) and $K = 2$ (14 parameters to estimate). In each case, 3 configurations are tested:

- **Scenario 1:** G trajectories starting from G different X_0^g , with n observations sampled at regular times ;
- **Scenario 2:** G trajectories starting from a unique X_0 , with n observations sampled at regular times (time step set at $\delta = 0.25$);
- **Scenario 3:** G trajectories starting from a unique X_0 , with n observations sampled at irregular times. Trajectories with $5n$ points are simulated with a regular time step $\delta = 0.05$, and n observations are drawn uniformly from these trajectories.

The first configuration is an "ideal case" as the exploration of the space is better with different starting points. The second and third configurations are more likely to happen in a real context since tagged individuals often start there trips from a unique point such as a colony or a harbor. Moreover, in ecology, most of timesteps acquisition of GPS positions are irregular due to environmental conditions. Examples of simulated trajectories for ($K = 2$, scenario 2) are shown in Figure 2. For $K = 1$ and $K = 2$, we set $G = 500$, $n = 31$, corresponding to numerous but short trajectories. For each sampling scheme, the MCEM algorithm is performed from 50 starting points $\theta_0^{(i)}, i = 1, \dots, 50$ giving 50 estimates $\hat{\theta}^{(i)}, i = 1, \dots, 50$. Then, the final estimate is given by

$$\hat{\theta} = \operatorname{argmax}_i \sum_{g=1}^G L^N(\mathbf{Y}^g, \hat{\theta}^{(i)}),$$

where $L^N(\mathbf{Y}^g, \hat{\theta}^{(i)})$ is the estimator of the loglikelihood $L(\mathbf{Y}^g, \hat{\theta}^{(i)})$ of the g -th trajectory (see Appendix C). The number of Monte Carlo samples to approximate the expectation increases with EM iterations (following [FM03]), up to a maximum of 100 particles per segment.

Choosing the starting point θ_0 of EM algorithm A key aspect in the behaviour of the EM algorithm deals with the initial set of parameters: choosing an appropriate θ_0 is crucial to design a time efficient estimation procedure. We consider some heuristics to pick a first guess, that would, of course, depend on the experiment.

- First guess for μ and C may be chosen using Gaussian mixture estimators, ignoring the correlation between successive relocations. However, this technique requires a subsampling of trajectories to get rid of the autocorrelation of the observed processes. This method may also be used to get the relative weight of each attractive zone.
- Choosing a good first guess for γ may be done using different estimators for diffusion process, for instance using Brownian bridges techniques. In practice, the estimation of this parameter is hardly sensitive to the starting point.
- A first guess for $\sum \pi_k$ may be trickier to find as it is strongly related to the speed of the individuals and thus requires expert knowledge of the experimental setting. It might be set to one.

Several values are drawn around this heuristically chosen starting set of parameters, corresponding to several trajectories of convergence. Finally the best estimate is chosen by computing the approximate log likelihood, as explained in appendix C.

Results An example of MCEM trajectories for ($K = 1$, scenario 3) is shown in Figure 3. Map estimates for $K = 1$ and $K = 2$ are presented in Figures 4 and 5. Detailed values for the best estimates are shown in Tables 1 and 2. As general comments, for all scenarios, there is a very fast convergence for the parameter γ . In these simple scenarios, our algorithm provides efficient estimates for the location parameters μ . The convergence is often slower for weights parameters π_k and shape parameters C_k . As expected, best estimates are obtained when the sampling is regular and the space is well explored (Scenario 1). Good estimation behavior is still observed when only one starting point is considered, and irregularity in sampling seems to have no impact on the performance of the estimation. This last point might be of great importance when dealing with actual data, as environmental conditions often lead to irregular sampling.

4.2 Real data set

GPS positions of a French fishing vessel performing in the English Channel were recorded during one year. The data set consists of 57 trajectories (assumed to be independent). Each trajectory is sampled regularly every 15 minutes in average, with about 40 points per trajectory, giving a total of 2723 points. The raw data set is shown on Figure 6a. Model (1) described above is fitted to this data set with $K = 1$, $K = 2$ and $K = 3$ modes⁵. Starting points for the EM algorithm were chosen randomly, using uniform distributions for position parameters (μ_k) weight parameters (π_k) and diffusion parameter γ , and Wishart distributions for shape parameters (C_k). A hundred iterations of the MCEM algorithm were performed for each point. For this data set, the model with three modes was selected using the approximated AIC criterion. The estimated maps for all values of K

⁵The model was coded and fitted using the **R** software [R C14] coupled with C++, using the package **Rcpp** [EF11]. All codes are available on demand.

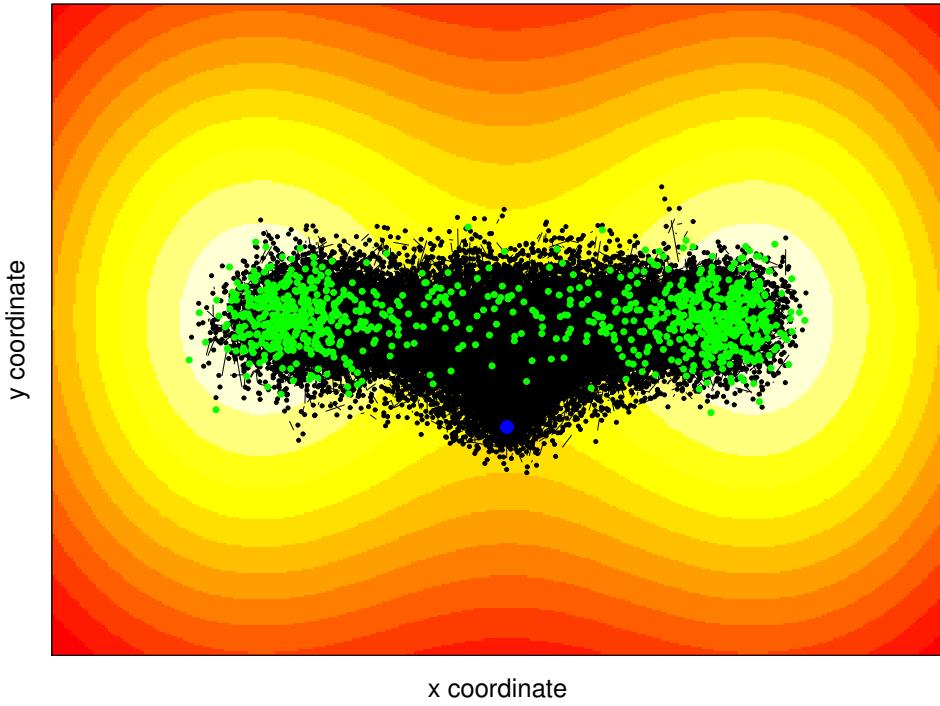


Figure 2: 1000 simulated trajectories for $K = 2$ starting from a unique X_0 (blue point). Each trajectory is sampled 30 times. Final points are colored in green. High (resp. low) potential areas of the underlying map are represented in white (resp. red).

Scenario/Parameter	π_1	$\mu^{(1)}$	$C^{(1)}$	γ
True Value	0.4	$(1, 0)'$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	0.2
Scen. 1	0.394	$(0.999, 0)'$	$\begin{pmatrix} 0.974 & 0.017 \\ 0.017 & 0.951 \end{pmatrix}$	0.196
Scen. 2	0.39	$(1.019, -0.002)'$	$\begin{pmatrix} 0.931 & -0.015 \\ -0.015 & 1.006 \end{pmatrix}$	0.196
Scen. 3	0.388	$(1.046, -0.008)'$	$\begin{pmatrix} 1.013 & -0.113 \\ -0.113 & 1.054 \end{pmatrix}$	0.197

Table 1: Best estimates for each scenario when $K = 1$. Results are rounded to 10^{-3} .

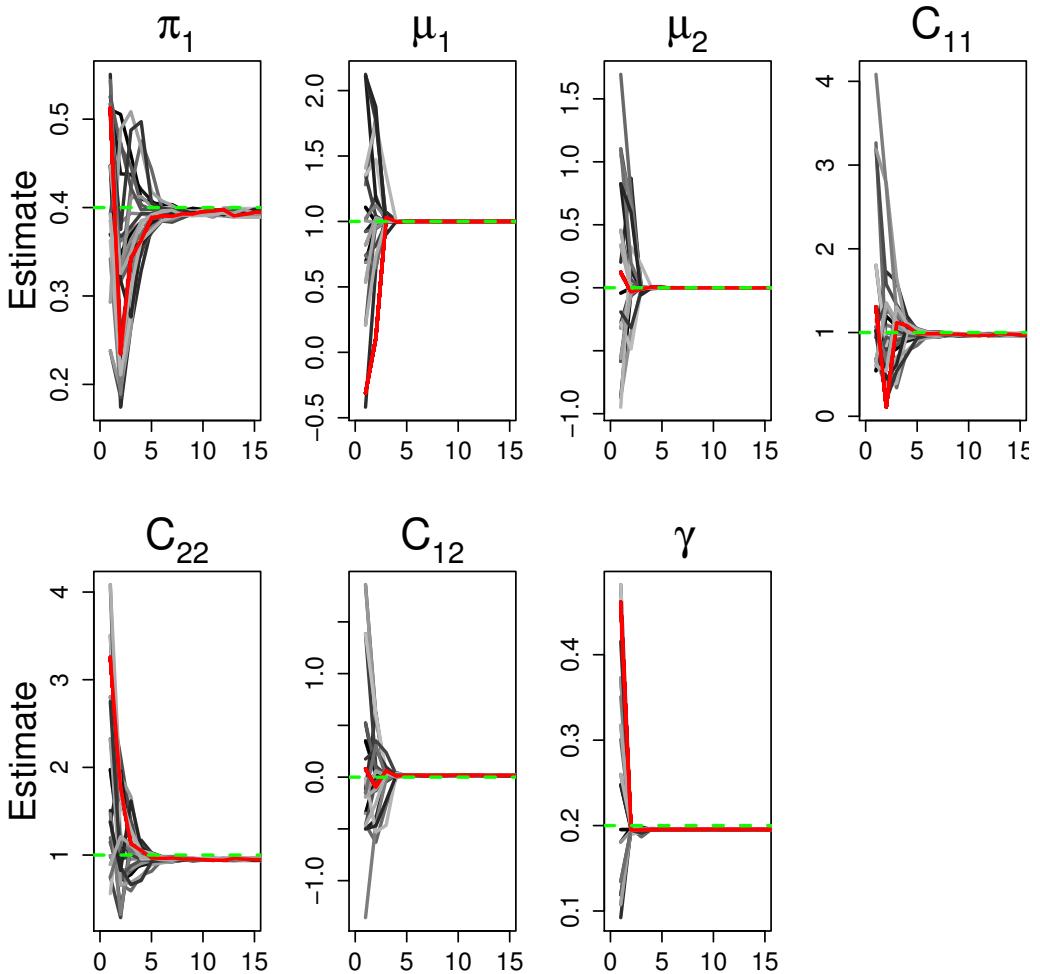


Figure 3: Example of MCEM trajectories for ($K = 1$, scenario 3). The red line is the best estimate and the green dotted line is the true value.

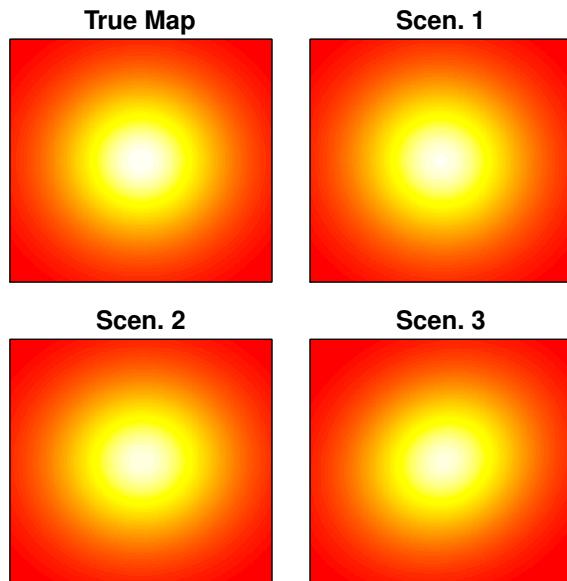


Figure 4: Estimated map for the three scenarios when $K = 1$. The scale is the same on each graph.

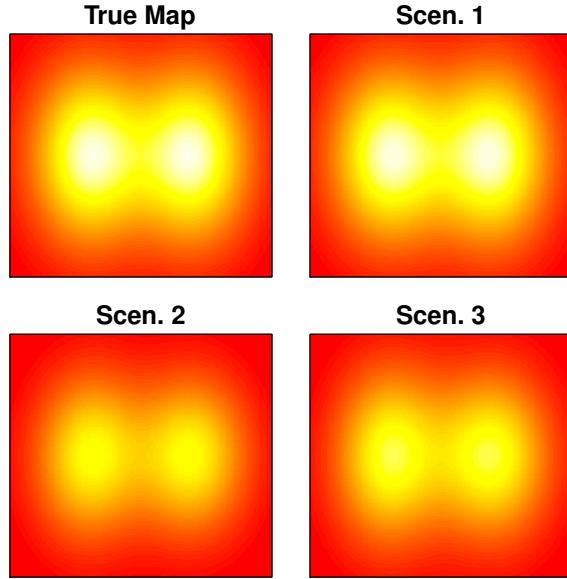


Figure 5: Estimated map for the three scenarios when $K = 2$. The scale is the same on each graph.

Scenario/Parameter	π_1, π_2	$\mu^{(1)}, \mu^{(2)}$
True Value	0.5, 0.5	$(1, 0)', (-1, 0)'$
Scen 1.	0.498, 0.497	$(1.01, 0)', (-1.01, 0.002)'$
Scen 2.	0.399, 0.399	$(0.969, -0.002)', (-0.981, -0.005)'$
Scen 3.	0.434, 0.430	$(0.974, -0.005)', (-0.991, -0.003)'$

Scenario/Parameter	$C^{(1)}, C^{(2)}$	γ
True Value	$\begin{pmatrix} 1.667 & 0 \\ 0 & 1.667 \end{pmatrix}, \begin{pmatrix} 1.667 & 0 \\ 0 & 1.667 \end{pmatrix}$	0.1
Scen 1.	$\begin{pmatrix} 1.644 & -0.012 \\ -0.012 & 1.633 \end{pmatrix}, \begin{pmatrix} 1.628 & -0.007 \\ -0.007 & 1.684 \end{pmatrix}$	0.098
Scen 2.	$\begin{pmatrix} 1.825 & -0.017 \\ -0.017 & 2.241 \end{pmatrix}, \begin{pmatrix} 1.943 & -0.055 \\ -0.055 & 2.059 \end{pmatrix}$	0.098
Scen 3.	$\begin{pmatrix} 1.691 & -0.002 \\ -0.002 & 2.004 \end{pmatrix}, \begin{pmatrix} 1.920 & -0.039 \\ -0.039 & 1.829 \end{pmatrix}$	0.097

Table 2: Best estimates for each scenario when $K = 2$. Results are rounded to 10^{-3} .

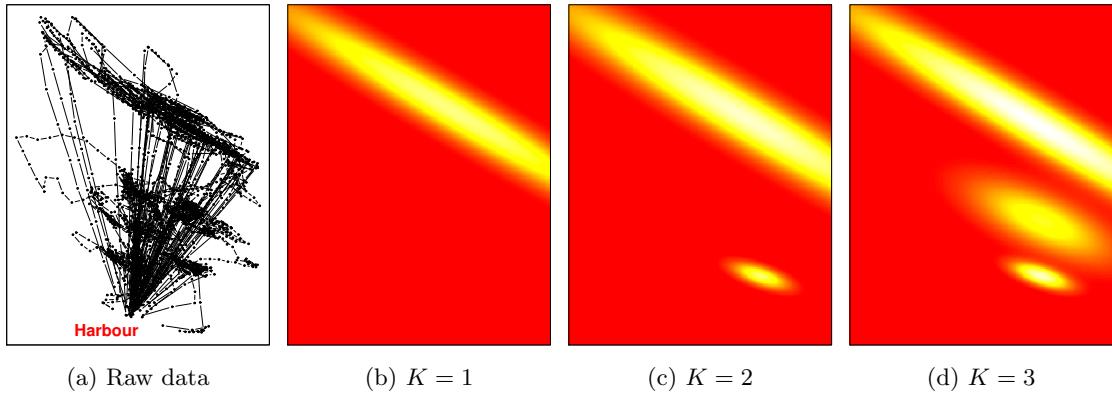


Figure 6: Estimated attractive zones for a French fishing vessel performing in the Channel. The model presented above is fitted with 1, 2 and 3 modes.

are given on figure 6. Three non connected zones are identified, a highly attractive deep sea zone, an attractive coastal zone, and an intermediate zone between two waters. The zones are oriented East-West, reflecting the main surface currents directions in the Channel.

5 Conclusions

This work proposes a new parametric model continuous in time and continuous in space to analyze trajectory data in ecology. This model assumes the trajectory process of an individual to be the realization of the solution to a SDE. The drift of this SDE reflects the attractiveness of the environment. More formally, the drift is defined as the gradient of a potential map. The model is estimated using an Monte Carlo EM algorithm. The Monte Carlo aspects rely on the exact algorithm for simulation proposed by [BPR06]. The exact simulation algorithm avoids the estimation bias due to discretization schemes, which may be large when the sampling frequency is not high enough. Here, the estimation of the potential map does not require any assumption on the sampling frequency, as would require kernel methods or discrete time models. The potential map is chosen here to be a mixture of positive Gaussian shaped functions. The parametric form of $P(\cdot)$ allows a flexible form for maps, however, the number of parameters increases linearly with the number of modes.

A minor change in this work would be to modify the parametric form of $P(\cdot)$ to describe other shapes for attractive areas, keeping only the boundary and C^2 properties. Such a change would only require a new computation of technical bounds required for the simulation procedure. As in any mixture model, the model selection problem arises and should be investigated in our framework. As an approximation of the loglikelihood is available, a penalized likelihood based criterion can be used. The approximation of the AIC criterion adopted here is the one proposed by [UY05] and [Iac09]. An important point here is that the criterion assumes the process solution to the SDE to be ergodic, as this criterion is mostly computed using one unique long realization. The process considered in equation (1) is not ergodic and exhibits Brownian Motion like behaviour when far from the attractive zone. Nevertheless good asymptotic properties are obtained as the number of trajectories G considered goes to infinity and the property of ergodicity is no more required. Since this model is intended to be used for ecological trajectory data sets and the structure of the data will often consist in a large amount of (presumably) independent trajectories that allows the state space to be visited. The ergodic property for the solution of our process would be satisfied at the cost of a change in the drift adding a term which avoids the process to visit space to far from the attractive points.

This model assumes that the diffusion coefficient is scalar, which is mainly a technical constraint. Using the approach presented here, the diffusion must satisfy a conservative property, that is actually the core of our model. A major limitation of our approach for ecological users might be the time homogeneous potential map. This assumption guarantees the process solution to the SDE (1) to have Markovian properties which greatly simplifies the estimation procedure. However, it is known that an individual might adopt different behaviors during its trips. It might be interesting to introduce state space models where different behaviors are considered.

Another interesting improvement would be to add interactions between individuals. An attraction/repulsion term could be added in the drift to indicate whether individuals are likely to

cooperate or not. Another limitation of our process is the non existence of a stationary distribution which is linked to the non ergodic property mentioned before. This is not a problem in practice since the starting point is close to attractive zones but, in theory, ecologists would rather consider an additional term to ensure the existence of a stationary distribution. This would be mandatory if considering only one unique long trajectory. As said above, if this term still allow the function in (7) to satisfy the conservative property, the estimation framework presented here would remain valid.

To conclude, we believe the model presented here offers new insight for trajectory analysis. We propose a general model continuous in time and space, that requires no assumption on sampling frequency. We think that this gradient based model and its estimation framework could be easily extended to answer many ecological questions.

References

- [AS08] Y. Aït-Shalia. Closed-form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36(2):906–937, 2008.
- [BB08] F. Barraquand and S. Benhamou. Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous bouts. *Ecology*, 89(12):3336–3348, 2008.
- [BCG00] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(7):719–725, Jul 2000.
- [BPR06] A. Beskos, O. Papaspiliopoulos, and G.O. Roberts. Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098, 2006.
- [BPR09] A. Beskos, O. Papaspiliopoulos, and G. Roberts. Monte Carlo maximum likelihood estimation for discretely observed diffusions processes. *Annals of Statistics*, 37(1):223–245, 2009.
- [BPRF06] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *Journal of Royal Statistical Society, Series B: Statistical Methodology*, 68:333–382, 2006.
- [BR05] A. Beskos and G.O. Roberts. Exact simulation of diffusions. *Annals of Applied Probability*, 15(4):2422–2444, 2005.
- [CFM⁺15] Fleming C.H., W.F. Fagan, T. Mueller, K.A. Olson, P. Leimgruber, and M.J. Calabrese. Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*, 96(5):1182–1188, 2015.
- [CG06] A.S. Chavez and E.M. Gese. Landscape Use and Movements of Wolves in Relation to Livestock in a Wildland–Agriculture Matrix. *Journal of Wildlife Management*, 70(4):1079–1086, 2006.
- [Cha11] S.-K. Chang. Application of a vessel monitoring system to advance sustainable fisheries management—Benefits received in Taiwan. *Marine Policy*, 35(2):116–121, 2011.
- [DG02] G.B. Durham and A.R. Gallant. Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes (with discussion). *J. Bus. Econom. Statist.*, 20:297–338, 2002.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. B*, 39(1):1–38 (with discussion), 1977.
- [ECS01] O. Elerian, S. Chib, and N. Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69:959–993, 2001.
- [EF11] D. Eddelbuettel and R. Franvois. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18, 2011.
- [Era01] B. Eraker. MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, 19:177–191, 2001.

- [FM03] G Fort and E Moulines. Convergence of the Monte Carlo expectation maximization for curved exponential families. *Annals of Statistics*, 31(4):1220–1259, 2003.
- [GMR⁺15] P. Gloaguen, S. Mahévas, E. Rivot, M. Woillez, J. Guitton, Y. Vermand, and M. P. Etienne. An autoregressive model to describe fishing vessel movement and activity. *Environmetrics*, 26(1):17–28, 2015.
- [HB13] K.J. Harris and P.G. Blackwell. Flexible continuous-time modelling for heterogeneous animal movement. *Ecological Modelling*, 255:29–37, 2013.
- [HK04] N. Hansen and S. Kern. Evaluating the CMA Evolution Strategy on Multimodal Test Functions. *Eighth International Conference on Parallel Problem Solving from Nature*, 72:337–354, 2004.
- [HO01] N. Hansen and A. Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evolutionary Computation*, 9(2):159–195, 2001.
- [Iac09] S.M. Iacus. *Simulation and Inference for Stochastic Differential Equations: With R Examples*. Springer Series in Statistics. Springer, 2009.
- [JBTF13] R. Joo, S. Bertrand, J. Tam, and R. Fablet. Hidden Markov Models: The Best Models for Forager Movements? *PLoS ONE*, 8(8):e71246, 08 2013.
- [JLLD08] D.S. Johnson, J.M. London, M.-A. Lea, and J.W. Durban. Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215, 2008.
- [PAJK04] H.K. Preisler, A.A. Ager, B.K. Johnson, and J.G. Kie. Modeling animal movements using stochastic differential equations. *Environmetrics*, 15:643–657, 2004.
- [PAW13] H.K. Preisler, A.A. Ager, and M.J. Wisdom. Analyzing animal movement patterns using potential functions. *Ecosphere*, 4(3), 2013.
- [Ped95] A.R. Pedersen. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, 1:257–279, 1995.
- [R C14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [RS01] G.O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88:603–621, 2001.
- [Sor04] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *Internat. Statist. Rev.*, 72:337–354, 2004.
- [UY05] M. Uchida and N. Yoshida. Aic for ergodic diffusion processes from discrete observations. *preprint MHF*, 12, 2005.

A Exact simulation of trajectories

This section details the mechanism to sample Y_u conditionally on (Y_0, Y_T) , where $(Y_t)_{0 \leq t \leq T}$ is a process solution to (6). Using the exact algorithm EA1 of [BPR06], this can be done using a rejection sampling algorithm based on Brownian bridges obtained at some random time steps. Let \mathbb{Q}_θ^y (resp. \mathbb{W}^y) be the probability measure induced by $(Y_t)_{0 \leq t \leq T}$ (resp. $(W_t)_{0 \leq t \leq T}$) on $(\mathcal{C}, \mathcal{C})$ conditioned on hitting y at time T . By (5), Girsanov theorem implies that

$$\frac{d\mathbb{Q}_\theta^y}{d\mathbb{W}^y}(\omega) \propto \exp \left\{ -\frac{1}{2} \int_0^T \left\{ \|\alpha_\theta(\omega_s)\|^2 + \Delta H_\theta(\omega_s) \right\} ds \right\}. \quad (12)$$

By lemmas B.1 and B.2, there exists $\ell_\theta \in \mathbb{R}$ such that for all $x \in \mathbb{R}^2$,

$$\frac{1}{2} \left(\|\alpha_\theta(x)\|^2 + \Delta H_\theta(x) \right) \geq \ell_\theta.$$

Then, if ϕ_θ is given by $\phi_\theta : x \mapsto \frac{1}{2} (\|\alpha_\theta(x)\|^2 + \Delta H_\theta(x)) - \ell_\theta$, the likelihood ratio (12) can be written:

$$\frac{d\mathbb{Q}_\theta^y}{d\mathbb{W}^y}(\omega) \propto \exp \left\{ - \int_0^T \phi_\theta(\omega_s) ds \right\} \leq 1.$$

To draw a path with distribution \mathbb{Q}_θ^y , a path $(\omega_s)_{0 \leq s \leq T}$ is sampled from \mathbb{W}^y and accepted with probability p_θ given by

$$p_\theta = \exp \left(- \int_0^T \phi_\theta(\omega_s) ds \right). \quad (13)$$

This acceptance probability cannot be evaluated in practice but this computation can be avoided by accepting a proposed path from \mathbb{W}^y with the realization of an event of probability p_θ . In [BPRF06], p_θ is interpreted as the probability of an event associated with an inhomogeneous Poisson process on $[0, T]$ with intensity $\phi_\theta(\omega_s)$. By lemmas B.1 and B.2, there exists Λ such that

$$\sup_{s \in [0, T]} \phi_\theta(\omega_s) \leq \Lambda_\theta.$$

By [BPRF06, Theorem 1], if Φ is a Poisson process on $[0, T] \times [0, \Lambda_\theta]$ with intensity $\Lambda_\theta T$ and if N_Φ is the number of points of Φ below the graph of $s \mapsto \phi_\theta(\omega_s)$, then

$$\mathbb{P}[N_\Phi = 0 | (\omega_s)_{0 \leq s \leq T}] = p_\theta.$$

The mechanism goes as follows. Let M be distributed according to a Poisson random variable with parameter $\Lambda_\theta T$, $\{\tau_1, \dots, \tau_M\}$ be uniformly distributed on $[0, T]$ and $\{\Phi_1, \dots, \Phi_M\}$ be uniformly distributed on $[0, \Lambda_\theta]$. Then, a proposed path $(\omega_s)_{0 \leq s \leq t}$ from \mathbb{W}^y (sampled using Brownian bridge dynamics) is accepted as a path from \mathbb{Q}_θ^y if $\mathbb{I} = 1$, where

$$\mathbb{I} := \prod_{i=1}^M \mathbf{1}_{\phi_\theta(\omega_{\tau_j}) < \Lambda_\theta \Phi_j}. \quad (14)$$

If the proposed path is accepted it can then be filled at time u using brownian bridge dynamics. This procedure is displayed in Algorithm 1.

Algorithm 1 Exact algorithm to sample Y_u conditionally on $(Y_0 = x, Y_T = y)$

- 1: **repeat**
 - 2: Let Λ_θ be such that $\sup_{x \in \mathbb{R}^2} \phi_\theta(x) \leq \Lambda_\theta$.
 - 3: Draw M , a Poisson random variable with parameter $\Lambda_\theta T$.
 - 4: Draw $\{\tau_1, \dots, \tau_M\}$ uniformly on $[0, T]$ and $\{\Phi_1, \dots, \Phi_M\}$ uniformly on $[0, \Lambda_\theta]$.
 - 5: Conditionally to $Y_0 = x$ and $Y_T = y$, draw a Brownian bridge at times $\{\tau_1, \dots, \tau_M\}$. The skeleton obtained is noted $\omega_{\tau_1} \dots \omega_{\tau_M}$
 - 6: Compute
 - $$\mathbb{I} = \prod_{i=1}^M \mathbf{1}_{\phi_\theta(\omega_{\tau_j}) < \Lambda_\theta \Phi_j} = 1.$$
 - 7: **until** $\mathbb{I} = 1$
 - 8: Find $k \in \{1 \dots M - 1\}$ such that $\tau_k \leq u < \tau_{k+1}$
 - 9: Draw Y_u from a Brownian Bridge, starting at ω_{τ_k} and ending at $\omega_{\tau_{k+1}}$.
-

B Implementation of EA1 for a mixture of Gaussian fields

Lemma B.1. *For all θ and all $x \in \mathbb{R}^2$,*

$$0 \leq \|\alpha_\theta(x)\|^2 \leq \bar{\alpha}_\theta,$$

where α_θ is given by (7) and, with $\lambda_1^{(k)}$ and $\lambda_2^{(k)}$ the eigenvalues of C_k , $1 \leq k \leq K$,

$$\bar{\alpha}_\theta := e^{-1} \gamma^{-2} \bar{\pi} \sum_{k=1}^K \pi_k \max_{1 \leq j \leq 2} \lambda_j^{(k)} \quad \text{and} \quad \bar{\pi} := \sum_{k=1}^K \pi_k. \quad (15)$$

Proof. For all θ and all $x \in \mathbb{R}^2$, by convexity of $\|\cdot\|^2$,

$$\begin{aligned}\|\alpha_\theta(x)\|^2 &= \left\| \sum_{k=1}^K \pi_k \varphi_k(\gamma x) \gamma^{-1} C_k(\gamma x - \mu_k) \right\|^2, \\ &\leq \bar{\pi} \sum_{k=1}^K \pi_k \varphi_k^2(\gamma x) \|\gamma^{-1} C_k(\gamma x - \mu_k)\|^2, \\ &\leq \bar{\pi} \sum_{k=1}^K \pi_k \gamma^{-2} \|C_k(\gamma x - \mu_k)\|^2 \exp(-(\gamma x - \mu_k)^T C_k(\gamma x - \mu_k)).\end{aligned}$$

Let Λ_k be defined by $C_k = P_k^{-1} \Lambda_k P_k$ where Λ_k is the diagonal matrix with diagonal given by $(\lambda_1^{(k)}, \lambda_2^{(k)})$ and where $P_k P_k^t = I_2$. If $z_k := P_k(\gamma x - \mu_k)$ then,

$$\|\alpha_\theta(x)\|^2 \leq \bar{\pi} \gamma^{-2} \sum_{k=1}^K \pi_k \underbrace{z_k^T \Lambda_k^2 z_k \exp(-[z_k^T \Lambda_k z_k])}_{f(z_k)},$$

where we used $\|P_k^T z_k\|^2 = \|z_k\|^2$ as P_k is orthogonal. The proof is concluded upon noting that for all $x \in \mathbb{R}^2$,

$$f(x) \leq e^{-1} \max_{1 \leq j \leq 2} \lambda_j^{(k)}.$$

□

Lemma B.2. For all θ and all $x \in \mathbb{R}^2$,

$$\Delta_\theta^- \leq \Delta H_\theta(x) \leq \Delta_\theta^+,$$

where ΔH_θ is given by (5) and, with $\lambda_1^{(k)}$ and $\lambda_2^{(k)}$ the eigenvalues of C_k , $1 \leq k \leq K$,

$$\Delta_\theta^- := - \sum_{k=1}^K \pi_k \text{Tr}(C_k), \quad (16)$$

$$\Delta_\theta^+ := 2e^{-1} \sum_{k=1}^K \pi_k \max_{1 \leq j \leq 2} \lambda_j^{(k)}. \quad (17)$$

Proof. By (5), if Tr denotes the Trace operator,

$$\begin{aligned}\Delta H_\theta(X) &= \text{Tr}[\nabla \alpha_\theta(x)], \\ &= -\text{Tr} \left[\nabla \left(\sum_{k=1}^K \pi_k \varphi_k(\gamma x) \gamma^{-1} C_k(\gamma x - \mu_k) \right) \right], \\ &= -\text{Tr} \left[\sum_{k=1}^K \left(\pi_k \nabla \varphi_k(\gamma x) \gamma^{-1} [C_k(\gamma x - \mu_k)]^T + \pi_k \varphi_k(\gamma x) \gamma^{-1} C_k \gamma \right) \right], \\ &= -\sum_{k=1}^K \pi_k \varphi_k(\gamma x) \left\{ -\text{Tr}([C_k(\gamma x - \mu_k)] [C_k(\gamma x - \mu_k)]^T) + \text{Tr}(C_k) \right\}, \\ &= \underbrace{\sum_{k=1}^K \pi_k \varphi_k(\gamma x) \|C_k(\gamma x - \mu_k)\|^2}_{I(x)} - \underbrace{\sum_{k=1}^K \pi_k \varphi_k(\gamma x) \text{Tr}(C_k)}_{J(x)}.\end{aligned}$$

By definition of φ_k , for all $1 \leq k \leq K$,

$$0 \leq J(x) \leq \underbrace{\sum_{k=1}^K \pi_k \text{Tr}(C_k)}_{-\Delta_\theta^-}.$$

Following the same steps as for the proof of Lemma B.1,

$$0 \leq I(x) \leq 2e^{-1} \underbrace{\sum_{k=1}^K \pi_k \max_{1 \leq j \leq 2} \lambda_j^{(k)}}_{\Delta_\theta^+}.$$

□

C Unbiased likelihood estimation for model selection

In the numerical Section, we need to estimate the loglikelihood $L(\mathbf{Y}^g, \theta)$ of each path $1 \leq g \leq G$ for a given parameter θ in order a) to choose the best estimate among the ones obtained from different starting points in the EM procedure and b) to compute the approximate AIC criterion for the chosen estimate to select the best number of components in our model. Note that,

$$L(\mathbf{Y}^g, \theta) = \sum_{j=0}^{n^g} \log p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta),$$

where $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$ is the conditional distribution of Y_j^g (at time t_j^g) given Y_{j-1}^g (at time t_{j-1}^g) when the parameter value is θ (with the convention that $p_{t_0^g - t_{-1}^g}(Y_0^g, Y_{-1}^g, \hat{\theta})$ is the likelihood of the first observation Y_0^g of path g). By [BPR09, Theorem 1], using the EA1 algorithm, it is possible, for any $1 \leq g \leq G$ and any $1 \leq j \leq n^g$, to define an unbiased estimator of $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$. Following the same steps as in [BPR09, Theorem 1], it can be proved that

$$\varphi_{t_j^g - t_{j-1}^g}(Y_j^g - Y_{j-1}^g) \exp \{H_\theta(Y_j^g) - H_\theta(Y_{j-1}^g) - \ell_\theta(t_j^g - t_{j-1}^g)\} \times \prod_{i=1}^{\Upsilon_\theta} (1 - \phi_\theta(\omega_{U_i})/\Lambda_\theta)$$

is an unbiased estimator of $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$ when ω has the law of a Brownian bridge between (t_{j-1}^g, Y_{j-1}^g) and (t_j^g, Y_j^g) , Υ_θ is a Poisson random variable with mean $\Lambda_\theta(t_j^g - t_{j-1}^g)$ and the $(U_i)_{1 \leq i \leq \Upsilon}$ are i.i.d. uniform random variables on $[0, t_j^g - t_{j-1}^g]$. Then, an unbiased Monte Carlo estimate of $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$ may be obtained by N independent realizations of these random variables. However, as the Poisson mean depends on θ , we would have to sample all the random variables for each value of the parameter θ . Nevertheless, in our case we need to estimate the likelihood only for a finite number of parameter values obtained at the end of the EM algorithm for each starting point: $\theta \in \{\hat{\theta}_i\}_{1 \leq i \leq p}$. In this case, writing $\bar{\Lambda} := \max_{1 \leq i \leq p} \Lambda_{\hat{\theta}_i}$, the unbiased estimator of $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$ we may use is given by

$$\varphi_{t_j^g - t_{j-1}^g}(Y_j^g - Y_{j-1}^g) \exp \{H_\theta(Y_j^g) - H_\theta(Y_{j-1}^g) - \ell_\theta(t_j^g - t_{j-1}^g)\} \times \prod_{i=1}^{\Upsilon} (1 - \phi_\theta(\omega_{U_i})/\bar{\Lambda}),$$

where ω has the law of a Brownian bridge between (t_{j-1}^g, Y_{j-1}^g) and (t_j^g, Y_j^g) , Υ is a Poisson random variable with mean $\bar{\Lambda}(t_j^g - t_{j-1}^g)$ and the $(U_i)_{1 \leq i \leq \Upsilon}$ are i.i.d. uniform random variables on $[0, t_j^g - t_{j-1}^g]$. As $\bar{\Lambda}$ is independent of θ , drawing once

- $(\omega^k)_{1 \leq k \leq N}$, N independent Brownian bridges between (t_{j-1}^g, Y_{j-1}^g) and (t_j^g, Y_j^g) ;
- $(\Upsilon_k)_{1 \leq k \leq N}$, N independent Poisson random variables with mean $\bar{\Lambda}(t_j^g - t_{j-1}^g)$;
- $\{(U_i^k)_{1 \leq i \leq \Upsilon}\}_{1 \leq k \leq N}$ independent uniform random variables on $[0, t_j^g - t_{j-1}^g]$;

allows to define the following estimator of $p_{t_j^g - t_{j-1}^g}(Y_j^g, Y_{j-1}^g, \theta)$ for all values $\theta \in \{\hat{\theta}_i\}_{1 \leq i \leq p}$:

$$\begin{aligned} p_{t_j^g - t_{j-1}^g}^N(Y_j^g, Y_{j-1}^g, \theta) &:= \varphi_{t_j^g - t_{j-1}^g}(Y_j^g - Y_{j-1}^g) \exp \{H_\theta(Y_j^g) - H_\theta(Y_{j-1}^g) - \ell_\theta(t_j^g - t_{j-1}^g)\} \\ &\quad \times \frac{1}{N} \sum_{k=1}^N \prod_{i=1}^{\Upsilon_k} (1 - \phi_\theta(\omega_{U_i}^k)/\bar{\Lambda}). \end{aligned}$$