

# A continuous time movement model linking displacement to utilization distribution

Pierre Gloaguen, Marie-Pierre Étienne and Sylvain Le Corff

January 5, 2018

## Abstract

This paper introduces a new continuous time and continuous space movement model, to describe animal's movement as the consequence of its environment quality and availability. The model depicts the position process as a diffusion process, solution to a Stochastic Differential Equation (SDE), with an explicit stationary distribution. This stationary distribution might be understood as the utilization distribution of the animal. The explicit mechanistic formulation through a stochastic differential equation allows the estimation of the utilization distribution using classical estimation methods for diffusion processes. This estimation allows the ecologist to test for the effect of any habitat preference on the utilization distribution, as well as to infer metrics of exploration tendency for a given individual.

## 1 Introduction

In movement ecology, a challenging question is to relate individuals movement to habitat preferences. As a proxy of an animal habitat preferences, ecologists often rely on the utilization distribution of the individual, that is the probability density function  $\pi$  such that, for an area  $\mathcal{A}$  of the space, if  $X \in \mathbb{R}^d$  denotes the position of the animal of interest:

$$\mathbb{P}(X \in \mathcal{A}) = \int_{\mathcal{A}} \pi(z) dz.$$

Environmental impact assessment and conservation measure require to predict the effect of a change on spatial environmental variables and therefore require knowledge on the link between environment variables and utilization distribution function. This link might be studied by incorporating  $K$  spatial covariates function  $c_1, \dots, c_K$ , in the utilization distribution function  $\pi$  and their effects summed up in a parameter  $\theta$  such that:

$$\pi(z) = \pi(\theta, c_1(z), \dots, c_K(z)). \quad (1)$$

In movement ecology, it is widely admitted that animal movement is related to the utilization distribution, as movement eventually leads to an overall space use. Indeed, movement data is often used to infer utilization distribution, either with kernel density methods (Worton, 1989; Fleming et al., 2015) or Brownian bridges methods (Horne et al., 2007). However, there exist few works to mechanistically link movement to the utilization distribution in a movement model. This would require to have a realistic movement model whose stationary distribution is the targeted utilization distribution. Recently Michelot et al. (2017) proposed a step selection movement model, explicitly based on the utilization distribution. This approach describes the individual's movement as a Markov chain whose stationary distribution is the utilization distribution. Markov Chain Monte Carlo (MCMC) algorithms may then be proposed to sample from the target stationary distribution. In this case, the authors used a local Gibbs sampler.

On the other hand, mechanistic movement model have been proposed where the individual's movement is a direct consequence of the environment (Brillinger, 2010). In this approach, the individual's position is supposed to follow the gradient of a potential surface whose value reflects the potential interest for the animal (Preisler et al., 2013). These Markovian approaches offer a wide variety of flexible models to describe movement, but their relation to utilization distribution is unclear in the existing examples as the diffusion process are not stationary (Gloaguen et al., 2017) or lead to unrealistically simple utilization distribution (by example a uniform distribution

for a Brownian motion movement model or a normal distribution for an Ornstein Uhlenbeck based movement model).

In this work, we propose a mechanistic movement model gathering the ideas of Brillinger (2010) (the movement is a diffusion process whose drift is the gradient of a potential surface) and of Michelot et al. (2017) (the stationary distribution of the animal's position is its utilization distribution). The movement model is based on the Langevin diffusion, known as the base of a particular MCMC algorithm (Roberts and Rosenthal, 1998).

As this model belongs to the class of potential based models, inference can be performed from movement data using different estimation methods for stochastic differential equations (SDEs), such as pseudo likelihood methods which are easy to implement (Gloaguen et al., 2017).

In section 2, we formulate the proposed movement model. In section 3, we briefly recall the inference method used in this work and present some results on simulated data.

## 2 Langevin movement model

We here consider an individual moving in its environment in  $\mathbb{R}^d$ ,  $d = 2$  mostly but might be equal to 3 in marine ecology, where the depth is of importance. We suppose that the individual has a utilization distribution  $\pi : \mathbb{R}^d \mapsto \mathbb{R}$ , potentially defined with a central parameter and the effect of some covariates, as in equation (1). We assume moreover that  $\pi(z)$  is strictly positive on  $\mathbb{R}^d$ , and that  $\log \pi(z)$  is smooth.

We suppose that the continuous time position process  $(X_t)_{t \geq 0}$  is the solution to the following stochastic differential equation:

$$dX_t = \frac{1}{2} \nabla_z \log \pi(X_t) dt + dW_t, \quad X_0 = x_0 \quad (2)$$

where  $W_t$  is a standard  $d$ -dimensional Brownian motion and  $x_0$  is the initial position of the animal. Under some mild technical conditions (that can be found in Dalalyan (2017)), the SDE (2) has a unique solution. This solution is a continuous time Markov process whose stationary distribution is  $\pi$  (Roberts and Tweedie, 1996).

### 2.1 Example

We here consider an example to show how the equation (2) can be used to simulate a trajectory from an individual having a known utilization distribution. We consider the following utilization distribution:

$$\pi(z) \propto \exp \left( \alpha \|z - \mu\|^2 + \beta_1 c_1(z) + \beta_2 c_2(z) \right) . \quad (3)$$

where:

- $c_1(z)$  and  $c_2(z)$  are smooth covariates whose gradient are assumed to be known for every  $z$ . A simple case, would be a linear effect of the covariates, but more complex effect can be modelled with some smooth transformation of the covariates. The example of covariates presented here are shown on Figure 1 ;
- $\beta_1$  and  $\beta_2$  are parameters indicating whether the following covariate has an attractive (positive sign) or repulsive (negative sign) effect on the individual's movement. For this example we set  $\beta_1 = -1$  and  $\beta_2 = 0.5$  ;
- $\mu$  is a home parameter, defining a point from which the animal tends to stay close. This point might reflect a nest or a burrow, or only the abstract center of a zone the animal does not quit for a long time. It is here fix to the center of the map (the 0 point) ;
- $\alpha$  is a concentration parameter around the point  $\mu$ . The higher alpha is, the closer the animal tends to stay to  $\mu$ . In the example, it is fixed to 0.05.

The resulting distribution given by (3) is shown in Figure 2.

**Simulations according to the Langevin model** The solution to equation (2) is a Markov process. However, for most cases, even with known parameters, there is no procedure to get sample exactly a solution to (2). Indeed, for a vast majority of diffusion processes, the probability density function of  $X_t$  given  $X_0$  has no explicit form, thus there is no straightforward exact sampling algorithm.

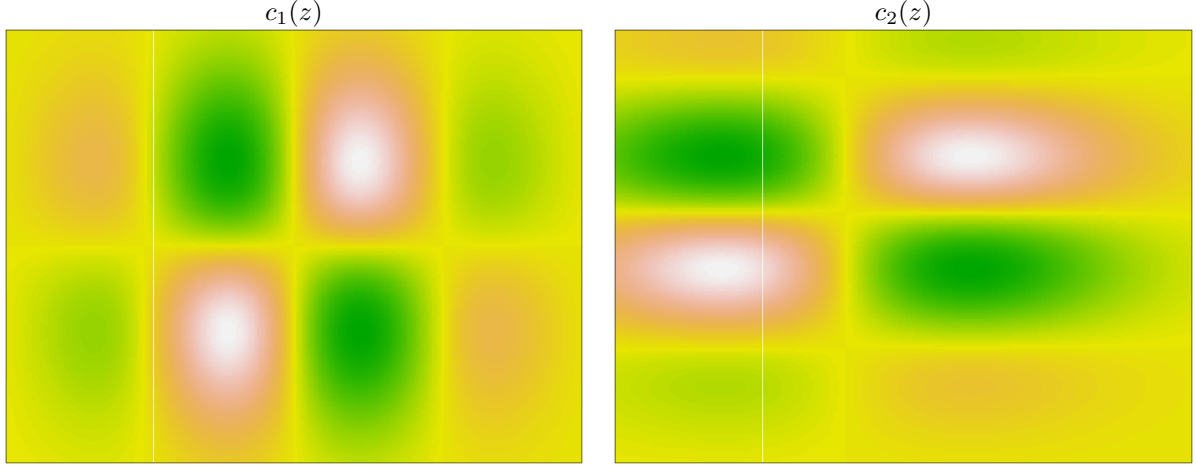


Figure 1: Two covariates patches for our example. Higher values are in white, lower values are in green. The covariate  $c_1$  is supposed to be repulsive, the covariate  $c_2$  is supposed to be attractive.

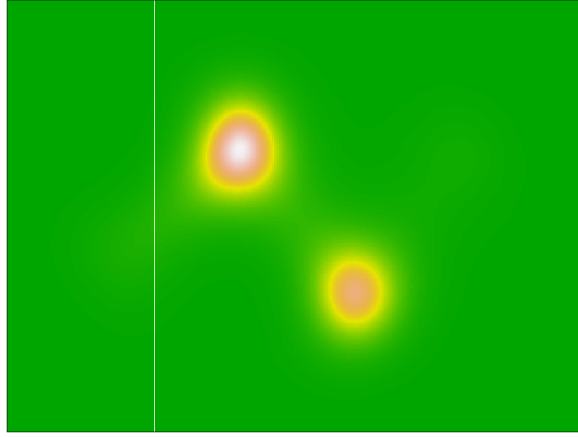


Figure 2: Utilization distribution for our example. This distribution is the result of equation (3) with parameters given in the text, and covariates of Figure 1.

However, some simple approximation schemes exist, such as the Euler scheme. This scheme can be used to sample approximations of the Langevin diffusion in the following way: from a starting point  $x_0$ , and given an approximation time step  $h$ , sample  $x_{t+h}$  as:

$$x_{t+h} = x_t + \nabla \log \pi(x_t) \times h + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}_d(0, h \times I_d) \quad (4)$$

This approximation is valid when  $h$  goes to 0. Indeed, it is strongly recommended to choose a small value for  $h$  so that the approximation remains valid (Roberts and Tweedie, 1996). An example of an approximated trajectory starting from  $\mu$  obtained using the Euler scheme is shown in Figure 3. More sophisticated approximate schemes might be explored (Gloaguen et al., 2017) with no restriction on the choice of the utilization distribution, while the exact sampling scheme (Beskos et al., 2006) imposes restrictive assumptions on  $\pi$ .

Despite the bimodal utilization distribution, the trajectory tends to explore the overall area and switches frequently from one mode to the other. Figure 4 shows the norm of  $\nabla \log \pi(z)$ , i.e. the drift of the SDE.

### 3 Inference

An important question in ecology is the inference of habitat preferences that shape the utilization distribution. In this context, given a parametric form for  $\pi(z)$  as in (1) or (3), the objective is to infer  $\theta$  from movement data. We suppose, for an individual, that data consists in a set of  $n$  points

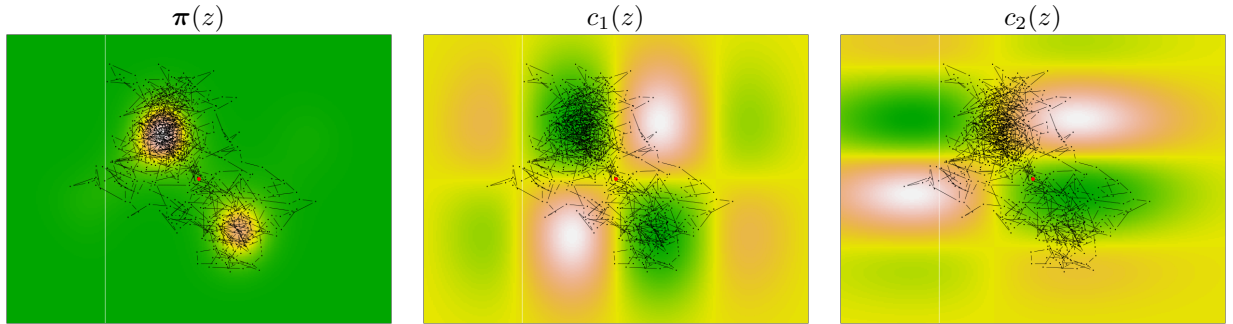


Figure 3: Example of a sampled trajectory using the Euler scheme  $n = 1000$ . The trajectory is plotted against the utilization distribution and the two covariates fields.

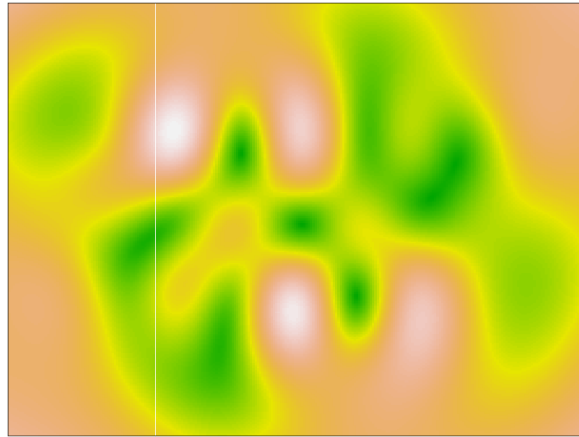


Figure 4: Gradient of  $\nabla \log \pi(z)$ . The greener, the lower, the whiter, the greater.

$\mathbf{x} := (x_1, \dots, x_n)$  sampled at times  $t_1 < t_2 < \dots < t_n$ . We suppose that this sample is a realisation of the process solution to (2). This solution is a Markov process, therefore, the likelihood is given by:

$$L(\theta|\mathbf{x}) = \sum_{i=1}^{n-1} p(x_{i+1}|x_i, \theta, \Delta_i), \quad (5)$$

where  $\Delta_i = t_{i+1} - t_i$ , and  $p(\cdot|x_i, \theta, \Delta_i)$  is the p.d.f. of the law of  $X_{t_i+\Delta_i}$  given that the process is at  $x_i$  at time  $t_i$  when the parameter is  $\theta$ .

However, as said above, this p.d.f. is impossible to compute in general. Different estimation methods exist to circumvent this problem (Gloaguen et al., 2017). Here, as a proof of concept, we choose to perform the estimation using the Euler-Maruyama approximation.

This method consists in replacing  $p(\cdot|x_i, \theta, \Delta_i)$  in equation (5) by the p.d.f. of a Gaussian distribution with mean  $x_i + \nabla \log \pi(z, \theta) \times \Delta_i$  and variance  $\Delta_i \times I_d$ , as suggested by (4)<sup>1</sup>.

### 3.1 Inference on simulated data

For any utilization distribution, the pseudo likelihood induced by the Euler approximation can be maximized using gradient ascent algorithm.

An example for the model considered here is given for a subsample of the simulations in Figure 3, with  $n = 500$  and  $\Delta = 1$ . Parameters  $\alpha$ ,  $\beta_1$  and  $\beta_2$  are estimated with 10 random starting points. The overall procedure takes less than one second. Results are shown in Figure 5. The algorithm converges after 3 iterations.

To assess whether, at this time step, the estimation is biased, the procedure is repeated for 100 similarly simulated trajectories and the associated boxplots are given in Figure 6. This shows an underestimation of parameter norms. This is expected at this time step, and might be accentuated by the Euler procedure.

## 4 Discussion

This preliminary work proposes a movement model linking movement to utilization distribution and habitat preferences. In this model, the path of an animal is considered as the solution to a Langevin diffusion. From this point of views, it presents strong analogies with the sampling scheme based on MCMC methods proposed by Michelot et al. (2017). As the model belongs to a known class of potential based diffusion processes, classical pseudo likelihood approaches can be used to estimate utilization distribution parameters from movement data.

This allows to test the influence of habitat covariates, as well as remoteness from a given point. On the example presented here (similar to the one of Michelot et al., 2017), parameter estimation is fast and explicit.

## References

- Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.
- Brillinger, D. (2010). *Handbook of Spatial Statistics*, chapter 26. Handbooks of Statistical Methods. Chapman and Hall/CRC Press.
- Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676.
- Fleming, C. H., Fagan, W. F., Mueller, T., Olson, K. A., Leimgruber, P., and Calabrese, J. M. (2015). Rigorous home range estimation with movement data: a new autocorrelated kernel density estimator. *Ecology*, 96(5):1182–1188.
- Gloaguen, P., Etienne, M.-P., and Corff, S. L. (2017). Stochastic differential equation based on a multimodal potential to model movement data in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

---

<sup>1</sup>This method is known to perform poorly when the sampling time step becomes too large, and could be replaced by other pseudo likelihood methods, such as the Ozaki method

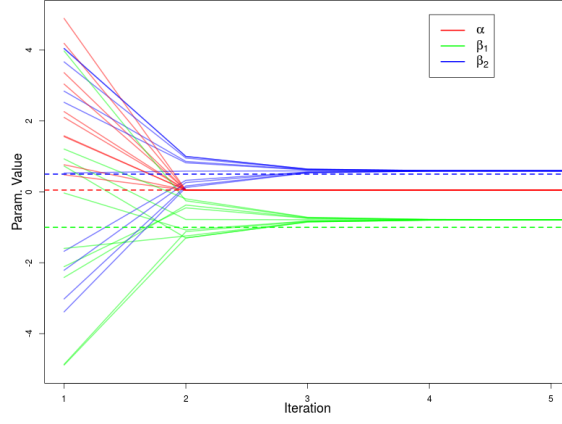


Figure 5: Estimations of  $\alpha$ ,  $\beta_1$  and  $\beta_2$  using a subsample of the trajectory of Figure 3, using 10 starting points for our estimation algorithm. True values for each parameter is the dotted line. The estimation converges after 3 iterations.

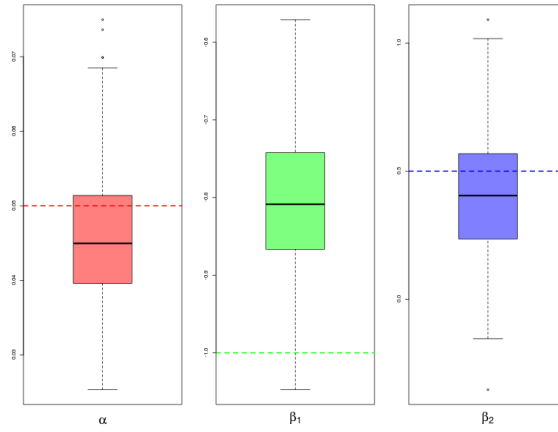


Figure 6: Boxplot of 300 independent estimates obtained on 300 independent trajectories with  $\Delta = 1$ . The bias is due to the large time step. It could also be due to the Euler approximation.

- Horne, J. S., Garton, E. O., Krone, S. M., and Lewis, J. S. (2007). Analyzing animal movements using brownian bridges. *Ecology*, 88(9):2354–2363.
- Michélot, T., Blackwell, P. G., and Matthiopoulos, J. (2017). Linking resource selection and step selection models for habitat preferences in animals. *arXiv preprint arXiv:1708.08426*.
- Preisler, H. K., Ager, A. A., and Wisdom, M. J. (2013). Analyzing animal movement patterns using potential functions. *Ecosphere*, 4(3):1–13.
- Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268.
- Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363.
- Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, 70(1):164–168.