



**HAL**  
open science

# The Langevin diffusion as a continuous-time model of animal movement and habitat selection

Théo Michelot, Marie-Pierre Etienne, Pierre Gloaguen

► **To cite this version:**

Théo Michelot, Marie-Pierre Etienne, Pierre Gloaguen. The Langevin diffusion as a continuous-time model of animal movement and habitat selection. 2018. hal-01900796v1

**HAL Id: hal-01900796**

**<https://hal.science/hal-01900796v1>**

Preprint submitted on 23 Oct 2018 (v1), last revised 28 Oct 2019 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

# The Langevin diffusion as a continuous-time model of animal movement and habitat selection

Théo Michelot<sup>1\*</sup>, Marie-Pierre Étienne<sup>2</sup>, Pierre Gloaguen<sup>3</sup>

<sup>1</sup>*University of Sheffield*, <sup>2</sup>*AgroCampus Ouest*, <sup>3</sup>*AgroParisTech*

## Abstract

1. The utilisation distribution describes the relative probability of use of a spatial unit by an animal. It is natural to think of it as the long-term consequence of the animal's short-term movement decisions: it is the accumulation of small displacements which, over time, gives rise to global patterns of space use. However, most utilisation distribution models either ignore the underlying movement, assuming the independence of observed locations, or are based on simplistic Brownian motion movement rules.
2. We introduce a new continuous-time model of animal movement, based on the Langevin diffusion. This stochastic process has an explicit stationary distribution, conceptually analogous to the idea of the utilisation distribution, and thus provides an intuitive framework to integrate movement and space use. We model the stationary (utilisation) distribution with a resource selection function to link the movement to spatial covariates, and allow inference into habitat selection.
3. Standard approximation techniques can be used to derive the pseudo-likelihood of the Langevin diffusion movement model, and to estimate habitat preference and movement parameters from tracking data. We investigate the performance of the method on simulated data, and discuss its sensitivity to the time scale of the sampling. We present an example of its application to tracking data of Stellar sea lions (*Eumetopias jubatus*).
4. Due to its continuous-time formulation, this method can be applied to irregular telemetry data. It provides a rigorous framework to estimate long-term habitat selection from correlated movement data.

**Keywords:** animal movement, continuous time, resource selection, step selection, Langevin diffusion, potential function, utilisation distribution

## 1 Introduction

A crucial concept in animal ecology is the utilisation distribution, "the probability density function that gives the probability of finding an animal at a particular location" (Anderson, 1982). In recent decades, improvements in tracking technologies have produced large amounts of animal location data, at a high spatio-temporal resolution. Statistical methods have been developed to estimate the UD from telemetry observations, and to link animal movement to habitat preferences and space use (Hooten et al., 2017). Most popular approaches to estimate the utilisation distribution from tracking data are non-parametric, and use empirical histograms (Nielsen and Sawyer, 2013) or kernel density estimation (Anderson, 1982; Worton,

---

\*Corresponding author: tmichelot1@sheffield.ac.uk

1989; Fleming et al., 2015). More recently, interpolation methods involving Brownian bridges have also been used (Horne et al., 2007; Kranstauber et al., 2012). A major limitation of such methods, however, is that the estimation of the utilisation distribution is disconnected from the movement itself, as they either ignore the sequential structure of data (Anderson, 1982; Worton, 1989; Nielson and Sawyer, 2013), or make unrealistic Brownian assumptions about the animal's movement (Horne et al., 2007; Kranstauber et al., 2012; Fleming et al., 2015).

In those approaches, a (generally two-dimensional) density function is estimated, and can be related to environmental covariates using regression techniques (Millspaugh et al., 2006; Long et al., 2009; Nielson and Sawyer, 2013; Zhang et al., 2014). The regression function, formulated in terms of spatial covariates of interest, is often defined as a resource selection function (RSF, Manly et al., 2007). A RSF links the distribution of observed locations of an animal to the distribution of resources (or other spatial covariates), to infer habitat characteristics that are preferred (or "selected") by the animal. It is based on the idea that, knowing the habitat composition of a spatial unit, we can predict its long-term utilisation. However, resource selection models usually assume that telemetry observations are independent, which is unrealistic for high-frequency movement data.

It is natural to think of the utilisation distribution as a consequence of the movement, which itself depends on the environment, such that short-term movement decisions give rise to long-term space use. This idea motivates the development of more mechanistic approaches that link the animal's movement to its environment, and, ultimately, a mechanistic movement model with an explicit steady-state distribution, representing the utilisation distribution.

Following this idea, step selection functions model the likelihood of a step between two points in space as a combination of a movement kernel and a habitat selection function (Fortin et al., 2005; Forester et al., 2009; Thurfjell et al., 2014). The parameters of a step selection function describe preference at a local (step-by-step) level, and strongly depend on the choice of the movement kernel. As such, it is unclear how they can be used to infer global space use. Potts et al. (2014) and Avgar et al. (2016) have suggested numerical methods to approximate the utilisation distribution underlying a step selection function model, but it does not take a parametric closed form.

Hanks et al. (2015) proposed a continuous-time discrete-space model to link movement to environmental drivers. In their framework, the movement is considered as a continuous-time Markov process on a discrete grid of spatial cells. The spatial grid is usually chosen as the grid on which the spatial covariates are measured, and the observed locations are binned in the cells. Wilson et al. (2018) argued that the limiting distribution of that movement model can be interpreted as the utilisation distribution of the animal, and proposed a method to estimate it on a discrete grid. A drawback of that approach is that it describes movement on a discrete spatial grid, and its formulation is therefore tied to a particular space discretization.

Recently, Michelot et al. (2017) proposed a step selection model, formulated in terms of an explicit utilisation distribution. Their approach describes individual movement as a Markov chain in continuous space, whose stationary distribution is the utilisation distribution. In particular, they suggest that Markov chain Monte Carlo (MCMC) algorithms, which are used to construct Markov chains with a given stationary distribution, can be viewed as movement models.

Others have described the position of an animal as a diffusion process which follows the gradient of a potential surface (Brillinger, 2010; Preisler et al., 2013; Gloaguen et al., 2018). The surface measures the potential interest for the animal, but it is not directly connected to the utilisation distribution. These approaches offer a wide variety of flexible models to describe movement, but their link to the utilisation distribution is unclear in the existing literature. Indeed, potential-based models are often based on diffusion processes that are not stationary (Gloaguen et al., 2018), or lead to unrealistically simple utilisation distributions.

For example, the stationary distribution is uniform over the study region for Brownian motion movement models (Skellam, 1951), and it is a normal distribution for Ornstein-Uhlenbeck based movement models (Blackwell, 1997).

In this work, we describe a new continuous-time and space mechanistic movement model. The animal's position is modelled as a diffusion process with a drift towards the gradient of its stationary (utilisation) distribution, bringing together the ideas of Brillinger (2010) and Michelot et al. (2017). As in Wilson et al. (2018), the limiting distribution of the process is the utilisation distribution. The movement model is based on the Langevin diffusion, which has also been used to construct a 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 simple to implement (Gloaguen et al., 2018). We show here how this parametric model can also be linked to step selection approaches when the utilisation distribution is parameterized as a simple function of environmental covariates. Point estimators and confidence intervals of habitat selection parameters can easily be derived in a classical approximated inference framework.

In Section 2, the proposed movement model is formulated in its general form, and in a specific "covariate-dependent" form, using a resource selection function. Section 3 describes a pseudo-likelihood method based on the Euler discretization scheme, for the estimation of habitat selection parameters from telemetry data. In Section 4, we assess the performance of inference methods in simulations, and we discuss conditions under which the model parameters can be recovered. In Section 5, we present the analysis of three trajectories of Stellar Sea Lions (*Eumetopias jubatus*) with the Langevin movement model, with four environmental covariates as potential drivers of their movement.

## 2 Langevin movement model

### 2.1 General formulation

We denote by  $\mathbf{X}_t \in \mathbb{R}^d$  the location of an animal in  $d$ -dimensional space at time  $t \geq 0$ , and  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}$  its utilisation distribution (Worton, 1989). The utilisation distribution is the probability density function such that, for any area  $A \subset \mathbb{R}^d$ ,

$$\mathbb{P}(\mathbf{X}_t \in A) = \int_A \pi(\mathbf{z})d\mathbf{z} \quad (1)$$

In the following, we will focus on the case  $d = 2$ , by far the most common in movement ecology, although the framework could be extended to higher dimensions.

We propose to describe the continuous-time location process of the animal  $(\mathbf{X}_t)_{t \geq 0}$  with a Langevin diffusion for the density  $\pi$ , defined as the solution to the stochastic differential equation

$$d\mathbf{X}_t = \frac{1}{2} \nabla \log \pi(\mathbf{X}_t) dt + d\mathbf{W}_t, \quad (2)$$

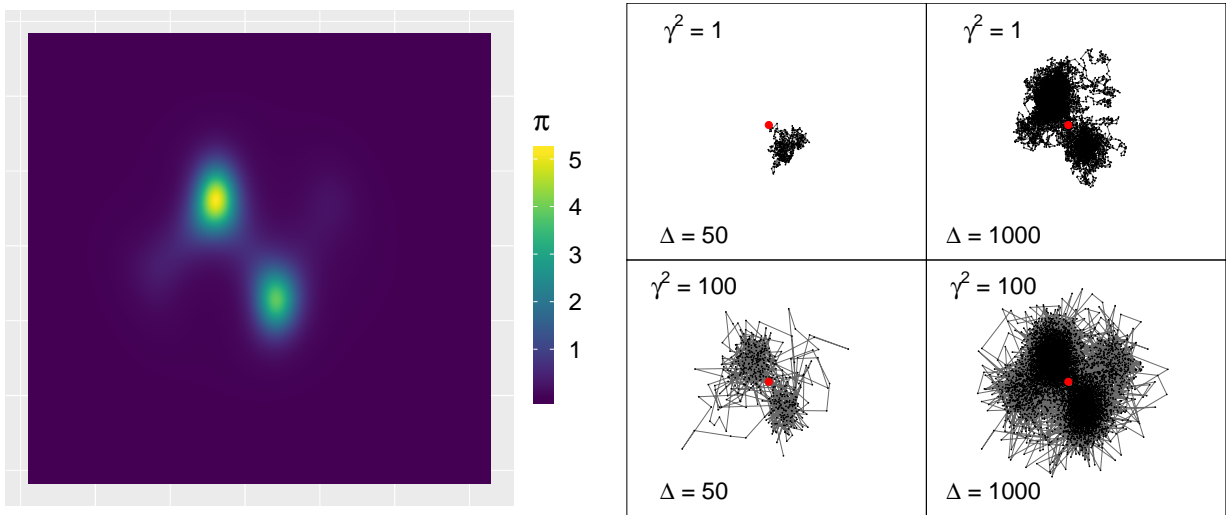
where  $\mathbf{W}_t$  is a standard Brownian motion,  $\nabla$  is the gradient operator, and with initial condition  $\mathbf{X}_0 = \mathbf{x}_0$ . Under some easily-satisfied technical conditions (that can be found in Dalalyan, 2017), Equation (2) has a unique solution, which is a continuous-time continuous Markov process. It describes the animal's movements as the combination of a drift towards higher values of its utilisation distribution  $\pi$  (informed by the gradient of  $\log \pi$ ), and a random component given by the Brownian motion. Crucially, the solution is a continuous-time Markov process whose asymptotic stationary distribution is  $\pi$  as defined in Equation (1) (Roberts et al., 1996). The Langevin diffusion is thus a natural choice for the basis of a continuous-time

model of animal movement, with a known steady-state distribution.

In its standard formulation, however, the Langevin diffusion cannot readily be used to model animal movement. Indeed, the speed of the process described above is only determined by the shape of the underlying utilisation distribution  $\pi$ , whereas it should be possible for two animals to move at different speeds on the same long-term distribution of space use. To allow for this flexibility, we introduce an additional parameter  $\gamma^2$ , and we define the Langevin movement model (with speed), as the solution to

$$d\mathbf{X}_t = \frac{\gamma^2}{2} \nabla \log \pi(\mathbf{X}_t) dt + \gamma d\mathbf{W}_t, \quad \mathbf{X}_0 = \mathbf{x}_0. \quad (3)$$

In the following,  $\gamma^2$  will be referred to as the speed parameter. The model is parameterized in terms of  $\gamma^2$  (rather than  $\gamma$ ) as it is also the variance parameter of the random Brownian motion component. Figure 1 shows two tracks simulated from the Langevin movement model on an artificial utilisation distribution, for two different values of  $\gamma^2$ . Although the two tracks explore space at very different speeds, they have the same equilibrium distribution.



**Figure 1:** Left: Artificial utilisation distribution  $\pi$ . Right: Trajectories simulated from the Langevin movement model on  $\pi$ , with two different values of the speed parameter  $\gamma^2$  (1 and 100), after  $\Delta = 50$  and  $\Delta = 1000$  time units. Although the process with  $\gamma^2 = 1$  is much slower to explore space, the properties of the Langevin equation guarantee that both processes have the same stationary distribution  $\pi$ .

## 2.2 Including covariates

We link the utilisation distribution of the animal to spatial covariates with the standard parametric form of resource selection functions (RSF),

$$\pi(\mathbf{x}|\boldsymbol{\beta}) = \frac{\exp\left(\sum_{j=1}^J \beta_j c_j(\mathbf{x})\right)}{\int_{\Omega} \exp\left(\sum_{j=1}^J \beta_j c_j(\mathbf{z})\right) d\mathbf{z}}, \quad (4)$$

where  $c_j(\mathbf{x})$  is the value of the  $j$ -th covariate at location  $\mathbf{x}$ ,  $\Omega \subset \mathbb{R}^d$  is the study region, and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$  is a vector of unknown parameters. Larger values of  $\beta_j$  indicate stronger selection of the  $j$ -th covariate. The denominator in the right-hand side of Equation (4) is a normalising constant, and is necessary to ensure

that  $\pi(\mathbf{x}|\boldsymbol{\beta})$  is a probability density function with respect to  $\mathbf{x}$ .

Note that Equation (3) requires  $\log \pi$  to be a smooth function, i.e. with continuous first-order partial derivatives. If  $\pi$  is modelled by a resource selection function (Equation (4)), then

$$\nabla \log \pi(\mathbf{x}|\boldsymbol{\beta}) = \sum_{j=1}^J \beta_j \nabla c_j(\mathbf{x}). \quad (5)$$

Therefore, it is supposed here that all covariates  $c_j$  are differentiable, and that their gradient can be computed at each point  $\mathbf{x}$ , either analytically, or by numerical approximation. In most real data sets, the covariate functions  $c_j$  are measured at discrete points in space. There is generally no analytical form for the gradient, and it is necessary to interpolate the covariate fields so that its gradient can be approximated. In Sections 4.2 and 5, bilinear interpolation is considered to obtain continuous covariate functions. As a consequence of this interpolation, the Langevin movement model cannot handle categorical covariates. Indeed, a categorical covariate field cannot be interpolated into a continuous function, and it is generally not possible to derive a measure of its gradient.

### 3 Inference

The continuous-time location process of the animal  $(\mathbf{X}_t)_{t \geq 0}$  is observed discretely at times  $t_0 < t_1 < \dots < t_n$ , and these observations are denoted by  $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n)$ . We consider  $J$  spatial covariates  $c_1, \dots, c_J$ , measured on a grid over the study region.  $\boldsymbol{\theta}$  denotes the vector of all parameters of the Langevin movement model defined in Section 2, i.e.  $\boldsymbol{\theta} = (\beta_1, \dots, \beta_J, \gamma^2)$ . This section describes an inference method to estimate  $\boldsymbol{\theta}$ , from telemetry and habitat data.

#### 3.1 Euler approximation of the likelihood

The likelihood of the observed locations, given  $\boldsymbol{\theta}$ , can be expressed using the *transition density* of the process  $(\mathbf{X}_t)_{t \geq 0}$ . The transition density is the probability density function of the random variable  $\mathbf{X}_{t+\Delta}$  given that  $\mathbf{X}_t = \mathbf{x}_t$ , and we denote it by  $q_\Delta(\mathbf{x}|\mathbf{x}_t, \boldsymbol{\theta})$ . By the Markov property of the solution to Equation (3), and assuming that the first position is deterministic, the likelihood function is

$$L(\boldsymbol{\theta}; \mathbf{x}_{0:n}) = \prod_{i=0}^{n-1} q_{\Delta_i}(\mathbf{x}_{i+1}|\mathbf{x}_i, \boldsymbol{\theta}), \quad (6)$$

where  $\mathbf{x}_{0:n}$  is shorthand for the set of observations, and  $\Delta_i = t_{i+1} - t_i$ .

As discussed in Gloaguen et al. (2018), in many practical cases, the density  $q_\Delta$  is intractable, and the likelihood  $L(\boldsymbol{\theta}; \mathbf{x}_{0:n})$  cannot be evaluated. To circumvent this problem, pseudo-likelihood approaches can be used as approximations. In these approaches, the intractable transition density in Equation (6) is replaced by the p.d.f. of a known distribution (usually, Gaussian), with moments given by a discretisation scheme.

The most common pseudo likelihood approach for discretely observed diffusion is the Euler discretization scheme (Iacus, 2009). In the Euler discretization (for  $d = 2$ ), the transition density of the Langevin diffusion is approximated by the following Gaussian density between  $t_i$  and  $t_{i+1}$ , for  $i = 0, \dots, n-1$ ,

$$\mathbf{X}_{i+1}|\{\mathbf{X}_i = \mathbf{x}_i\} = \mathbf{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\mathbf{x}_i|\boldsymbol{\beta}) + \varepsilon_{i+1}, \quad \varepsilon_{i+1} \stackrel{ind}{\sim} N(\mathbf{0}, \gamma^2 \Delta_i \mathbf{I}_2), \quad (7)$$

where  $\mathbf{I}_2$  is the  $2 \times 2$  identity matrix. Under this approximation, the transition density of the process can

then be written

$$q_{\Delta_i}(\mathbf{x}_{i+1}|\mathbf{x}_i, \boldsymbol{\theta}) = \phi\left(\mathbf{x}_{i+1}\left|\mathbf{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\mathbf{x}_i|\boldsymbol{\beta}); \gamma^2 \Delta_i \mathbf{I}_2\right.\right),$$

where  $\phi(\cdot|\boldsymbol{\mu}; \boldsymbol{\Sigma})$  is the p.d.f. of the multivariate normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . This expression can be plugged into Equation (6) to obtain the likelihood of a track  $\mathbf{x}_{0:n}$ .

The Euler discretization can also be used to simulate (approximately) from the Langevin movement model, as illustrated in the simulations of Section 4. The quality of the scheme decreases for longer time steps of simulation (Kessler et al., 2012, Chapter 1).

### 3.2 Maximum likelihood estimation

The pseudo-likelihood function could be optimised numerically to obtain estimates of all model parameters. However, if  $\pi$  is modelled with the resource selection function of Equation (4), the maximum likelihood estimate  $\hat{\boldsymbol{\theta}}$  can simply be obtained using standard linear model equations.

Plugging Equation (5) into Equation (7), we can write a standard linear model, in the following matrix form. Let  $\mathbf{Y}_i = (\mathbf{X}_{i+1} - \mathbf{X}_i)/\sqrt{\Delta_i}$  be the (two-dimensional) normalized random increment of the process between  $t_i$  and  $t_{i+1}$ , and denote

$$\mathbf{Y} = \begin{pmatrix} Y_{0,1} \\ \vdots \\ Y_{n-1,1} \\ Y_{0,2} \\ \vdots \\ Y_{n-1,2} \end{pmatrix}, \quad \mathbf{D} = \frac{1}{2} \begin{pmatrix} \frac{\partial c_1(\mathbf{x}_0)}{\partial z_1} & \frac{\partial c_2(\mathbf{x}_0)}{\partial z_1} & \cdots & \frac{\partial c_J(\mathbf{x}_0)}{\partial z_1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial c_1(\mathbf{x}_{n-1})}{\partial z_1} & \frac{\partial c_2(\mathbf{x}_{n-1})}{\partial z_1} & \cdots & \frac{\partial c_J(\mathbf{x}_{n-1})}{\partial z_1} \\ \frac{\partial c_1(\mathbf{x}_0)}{\partial z_2} & \frac{\partial c_2(\mathbf{x}_0)}{\partial z_2} & \cdots & \frac{\partial c_J(\mathbf{x}_0)}{\partial z_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial c_1(\mathbf{x}_{n-1})}{\partial z_2} & \frac{\partial c_2(\mathbf{x}_{n-1})}{\partial z_2} & \cdots & \frac{\partial c_J(\mathbf{x}_{n-1})}{\partial z_2} \end{pmatrix},$$

where  $\mathbf{Y}_i = (Y_{i,1}, Y_{i,2})$ , and  $\partial/\partial z_i$  denotes the partial derivative with respect to the  $i$ -th spatial coordinate.

Moreover, let  $\mathbf{T}_\Delta$  be the  $(2n) \times (2n)$  diagonal matrix with  $i$ -th and  $(n+i)$ -th diagonal terms equal to  $\sqrt{\Delta_{i-1}}$ , for  $i = 1, \dots, n$ . Then, the Euler approximation of the Langevin movement model can be rewritten as

$$\mathbf{Y} = (\mathbf{T}_\Delta \mathbf{D}) \boldsymbol{\nu} + \mathbf{E}, \quad (8)$$

where  $\mathbf{E}$  is a  $2n$ -vector of  $N(0, \gamma^2)$  variables, and  $\boldsymbol{\nu} = \gamma^2 \boldsymbol{\beta}$ . The estimators for  $\boldsymbol{\nu}$  and  $\gamma^2$  are derived from standard linear model theory, as

$$\hat{\boldsymbol{\nu}} = (\mathbf{D}^\top \mathbf{T}_\Delta^2 \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{T}_\Delta \mathbf{Y},$$

and

$$\hat{\gamma}^2 = \frac{1}{2n - J} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2,$$

where  $\hat{\mathbf{Y}} = (\mathbf{T}_\Delta \mathbf{D}) \hat{\boldsymbol{\nu}}$  is the predicted value of  $\mathbf{Y}$ . In Appendix A, we show that an unbiased estimator of  $\boldsymbol{\beta}$  is then given by

$$\hat{\boldsymbol{\beta}} = \frac{(2n - J - 2) \hat{\boldsymbol{\nu}}}{(2n - J) \hat{\gamma}^2},$$

with the following covariance formula

$$\text{Cov}(\hat{\beta}_j, \hat{\beta}_k) = \frac{2\beta_j \beta_k}{2n - J - 4} + \frac{\Upsilon_{jk}}{\gamma^2} \left(1 + \frac{2}{2n - J - 4}\right),$$

where  $\Upsilon_{jk} := [(D^\top T_\Delta^2 D)^{-1}]_{jk}$ . Using the asymptotic normality of maximum likelihood estimators, we can obtain a confidence interval of any level  $\alpha \in (0, 0.5)$ , that is valid for a large enough  $n$ . For each covariate coefficient  $\beta_j$ ,

$$CI_\alpha(\beta_j) = \left[ \hat{\beta}_j + z_{\alpha/2} \times \sqrt{\mathbb{V}(\hat{\beta}_j)}; \hat{\beta}_j - z_{\alpha/2} \times \sqrt{\mathbb{V}(\hat{\beta}_j)} \right]$$

where  $z_{\alpha/2}$  is the quantile of level  $\alpha/2$  of a standard Gaussian distribution, and  $\mathbb{V}(\hat{\beta}_j) = \text{Cov}(\hat{\beta}_j, \hat{\beta}_j)$ . The detail of the derivation is given in Appendix A.

The results above derive from the use of a linear model to approximate the solution to a SDE.

For the Langevin movement model based on a RSF, as defined in Section 2, the Euler approximation therefore provides explicit estimates and confidence intervals. Note that the Euler estimator is *biased* due to the approximation made in Equation (7) (see Kessler et al., 2012, Chapter 1). Therefore, both the estimate and the confidence interval must be interpreted with caution, as they depend on the quality of the scheme. The potential use of other discretization schemes is discussed in Section 6.

## 4 Simulation study

In this section, we assess the performance of the inference method described in Section 3 in two simulation scenarios. In both cases, we simulate movement tracks from the Langevin process, using the Euler discretization given in Equation 7. We simulate covariates and define an artificial utilisation distribution, expressed as a resource selection function, as shown in Equation (4). The objective is to recover the habitat selection parameters  $\{\beta_1, \dots, \beta_J\}$  and the speed parameter  $\gamma^2$ .

### 4.1 Scenario 1

We first consider a fully controlled simulation scenario, where the covariate fields are given by smooth analytical functions. In this idealized case, the gradient of the covariate functions, and thus of the utilisation distribution, can be calculated exactly at any point in the plane. The utilisation distribution  $\pi$  is defined as a RSF (Equation (4)) of three covariates  $c_1, c_2$  and  $c_3$ , given by

$$\begin{aligned} c_j(\mathbf{z}) &= \alpha_j \exp(-(\mathbf{z} - \mathbf{a}^j)^\top \Sigma^j (\mathbf{z} - \mathbf{a}^j)) \times \sin\left(\omega_1^j(z_1 - a_1^j)\right) \times \sin\left(\omega_2^j(z_1 - a_2^j)\right), \quad j = 1, 2, \\ c_3(\mathbf{z}) &= \|\mathbf{z}\|^2, \end{aligned}$$

where  $\alpha_j, \mathbf{a}^j = (a_1^j, a_2^j)$ ,  $\omega^j = (\omega_1^j, \omega_2^j)$ , and  $\Sigma^j = \text{diag}\{\sigma_1^j, \sigma_2^j\}$  are known simulation parameters whose values are given in Appendix B. For the simulations, we choose the resource selection parameters  $\beta_1 = -1$ ,  $\beta_2 = 0.5$ , and  $\beta_3 = -0.05$ , and the speed parameter  $\gamma^2 = 1$ .

The first two covariates are smooth functions, for which the gradient can easily be derived. The third covariate is the squared distance to the centre of the map, and is used to include a weak force of attraction towards the centre (here, the point  $(0, 0)$ , somewhat related to the home range of the individual). These three covariates functions are shown in Figure 2.

Inference was performed independently on 600 data sets. Each data set was a trajectory of 300 points, simulated from the Langevin movement model. The tracks were first generated at a fine time resolution ( $\Delta = 0.01$ ), to minimise the effect of the Euler approximation, and they were then thinned to time intervals of 0.5 time units.



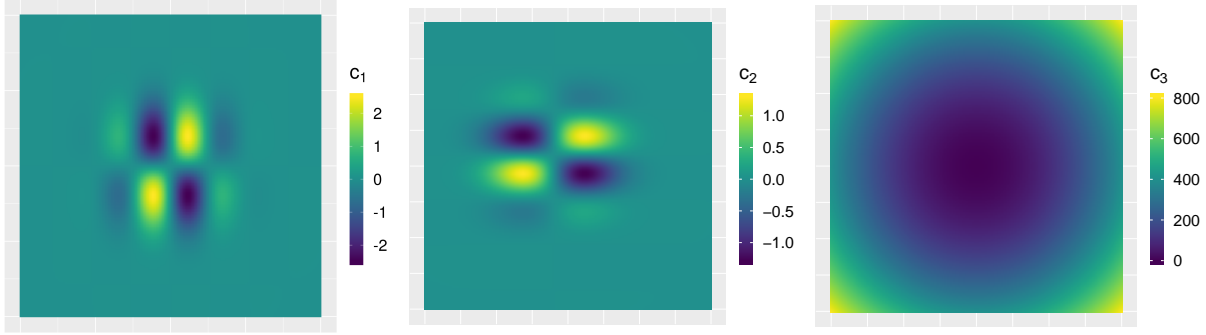


Figure 2: Artificial covariates fields for the simulation scenario of Section 4.1.

We estimated all model parameters using the Euler method, presented in Section 3.2. We considered two different settings: (i) the true analytic gradient is used in the estimation, and (ii) the covariates are discretized on a  $8 \times 8$  regular grid, and the gradient is obtained through the interpolation of the covariates. This second setting corresponds to the more realistic case where covariates are only observed on a discrete grid, and the gradient needs to be approximated. The gradient approximations were performed for the covariates  $c_1$  and  $c_2$  using the R package `nloptr` (Ypma et al., 2014). The gradient of the Euclidean distance  $c_3$  is computed exactly in both cases, as it would be in a real analysis.

Boxplots of the parameter estimates in the 600 replications are shown in Figure 3. All parameters were correctly estimated in this benchmark scenario, even when the covariates were discretized to a coarse grid.

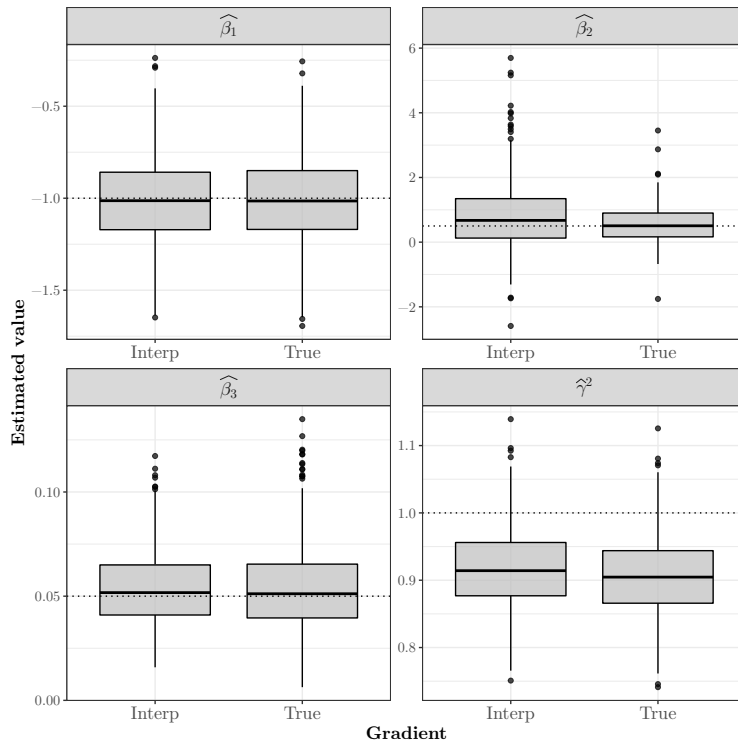
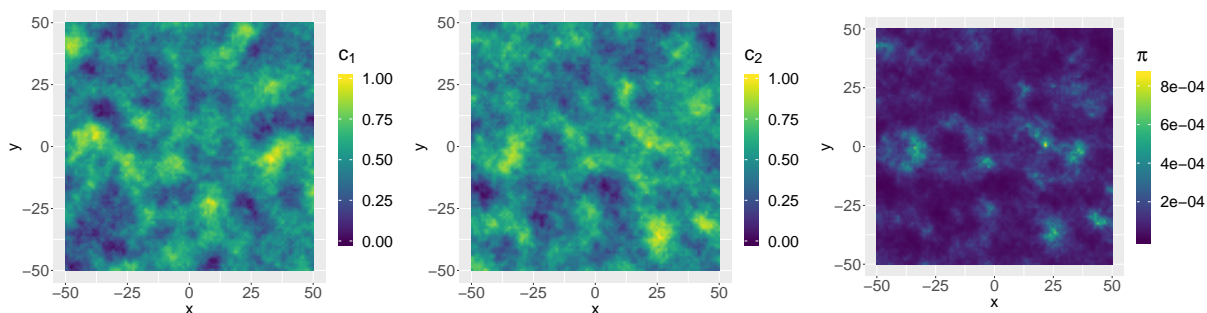


Figure 3: Estimates for model parameters on 600 experiments replications of scenario 1. The dotted lines show the real values used in the simulations.

## 4.2 Scenario 2

We considered a second simulation scenario, with randomly-generated covariate fields on a discrete grid, more similar to real environmental data. The main objective of this scenario is to investigate the effect of the sampling frequency on the estimation.

To simulate covariates, we used a procedure similar to that described by Avgar et al. (2016). We defined a spatial grid over  $[-50, 50] \times [-50, 50]$ , with cells of size 1. For each cell, a random uniform value was generated on  $(0, 1)$ , and the covariate field was obtained with a two-dimensional moving average filter over a circular region of radius  $\rho$ . Here,  $\rho$  measures the degree of spatial autocorrelation of the simulated covariate. We then normalized the covariate field, to range between 0 and 1. Using this procedure, we simulated two covariates  $c_1$  and  $c_2$ , with the same autocorrelation parameter  $\rho = 10$ . Then, we simulated trajectories from the Euler scheme described in Equation (7), with  $\pi$  defined as the (normalized) RSF with coefficients  $(\beta_1, \beta_2) = (2, 4)'$ . Plots of the simulated covariates, and of the utilisation distribution used in the simulations, are shown in Figure 4.

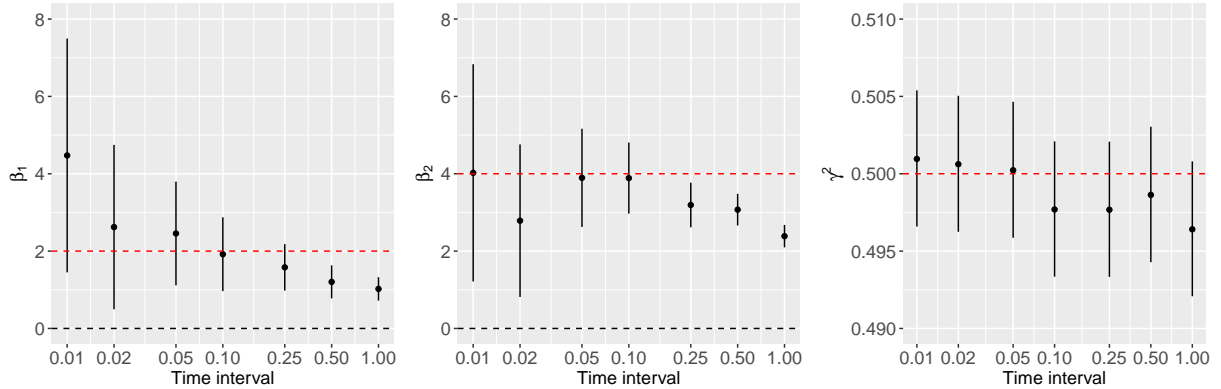


**Figure 4:** Simulated covariate fields  $c_1$  and  $c_2$ , and utilisation distribution obtained with  $\beta = (2, 4)'$ . used in the second simulation scenario.

We simulated 200 trajectories from the Langevin movement model, at a fine temporal resolution of  $\Delta = 0.01$ . We then subsampled each trajectory, for different time resolutions  $\Delta \in \{0.01, 0.02, 0.05, 0.1, 0.25, 0.5, 1\}$ , to emulate data sets obtained at different observation rates. From each thinned data set, we kept the first 250 locations of each of the 200 trajectories, leading to a total of 50,000 locations.

We fitted the Langevin movement model to each thinned data set, using the estimators given in Section 3.2. We evaluated the gradients of the covariates at each simulated location with a bilinear interpolation method. This procedure was done using the function `interp.surface` from the R package `fields` (Douglas Nychka et al., 2017). We then obtained the approximated gradient of the interpolated covariates with the R package `nloptr`. Point estimates and 95% confidence intervals of the habitat selection parameters  $\beta$  and the speed parameter  $\gamma^2$  are displayed in Figure 5.

The estimates of the speed parameter  $\gamma^2$  were very close to the true value, for all simulation experiments. However, there was a lot of variability in the accuracy and precision of habitat selection parameter estimates. The uncertainty on the estimates of the habitat selection parameters decreased as the time interval increased. This is not surprising: all trajectories had the same number of locations, such that those with longer time intervals explored a larger proportion of the study region. Tracks with longer time intervals therefore covered a larger range of covariate values. Similarly to standard linear model analyses, the uncertainty on the coefficients is larger when the observed range of explanatory variables in Equation (8) is narrow. From  $\Delta = 0.05$  to  $\Delta = 1$ , the estimates of  $\beta_1$  and  $\beta_2$  both decreased as the time interval increases, leading to an underestimation of the parameters for longer time intervals. This is a common problem for the estimation



**Figure 5:** Estimates of the habitat selection parameters  $\beta$  and of the speed parameter  $\gamma^2$  for different observation intervals, in the simulation study of Section 4.2. The vertical lines show 95% confidence intervals, and the red dotted lines are the true values of the parameters. The x axis is on the log scale.

of discretely observed diffusion processes, because the consistency of the estimators requires  $\Delta$  to tend towards 0 (for more details, see Kessler et al., 2012). For long time intervals, the habitat selection parameters are underestimated in absolute value, i.e. the strength of the (positive or negative) effect is underestimated. This bias is a side effect of the underestimation of the speed of the process for long time lags between observations. As the time interval increases, the estimated utilisation distribution becomes flatter, to reflect our growing uncertainty about the effect of the covariates on the short-term movement. In the extreme, for very long time intervals, we would have no information about the selection process, and the estimated utilisation distribution would be flat, corresponding to a uniform distribution of space use over the study region. Note that, although the strength of selection was underestimated in the simulations with long time intervals, the sign of the effect – i.e. selection or avoidance – was always estimated correctly (Figure 5).

To investigate the performance of the method for the analysis of data sets collected at irregular time intervals, we ran a similar experiment where the observations were thinned at random. The results were very similar to the simulations with regular intervals, and are presented in Appendix C. These findings confirm that, due to its continuous-time formulation, the Langevin movement model can directly be used on tracking data collected irregularly.

## 5 Example analysis

In this section, we fit the Langevin movement model to a data set described by Wilson et al. (2018), collected on Steller sea lions (*Eumetopias jubatus*) in Alaska. The data set comprises three trajectories, obtained from three different individuals, for a total of 2672 Argos locations. The time intervals were highly irregular, with percentiles  $P_{0.025} = 6\text{min}$ ,  $P_{0.5} = 1.28\text{h}$ ,  $P_{0.975} = 17.4\text{h}$ . In addition to the locations, Wilson et al. (2018) provided four spatial covariates over the study region, at a resolution of 1km: bathymetry ( $c_1$ ), slope ( $c_2$ ), distance to sites of interest ( $c_3$ ), and distance to continental shelf ( $c_4$ ). The sites of interest were either haul-out or rookery sites. Maps of the covariates are shown in Figure 6, and we refer the readers to Wilson et al. (2018) for more detail about the data set.

To correct for the measurement error, and to follow the preprocessing performed by Wilson et al. (2018), we first fitted a continuous-time correlated random walk (CTCRW) to the tracks, using the R package *crawl* (Johnson et al., 2008; Johnson and London, 2018). The CTCRW is a continuous-time state-space model, that can be used on irregular and noisy telemetry data. The package *crawl* implements the Kalman filter for this

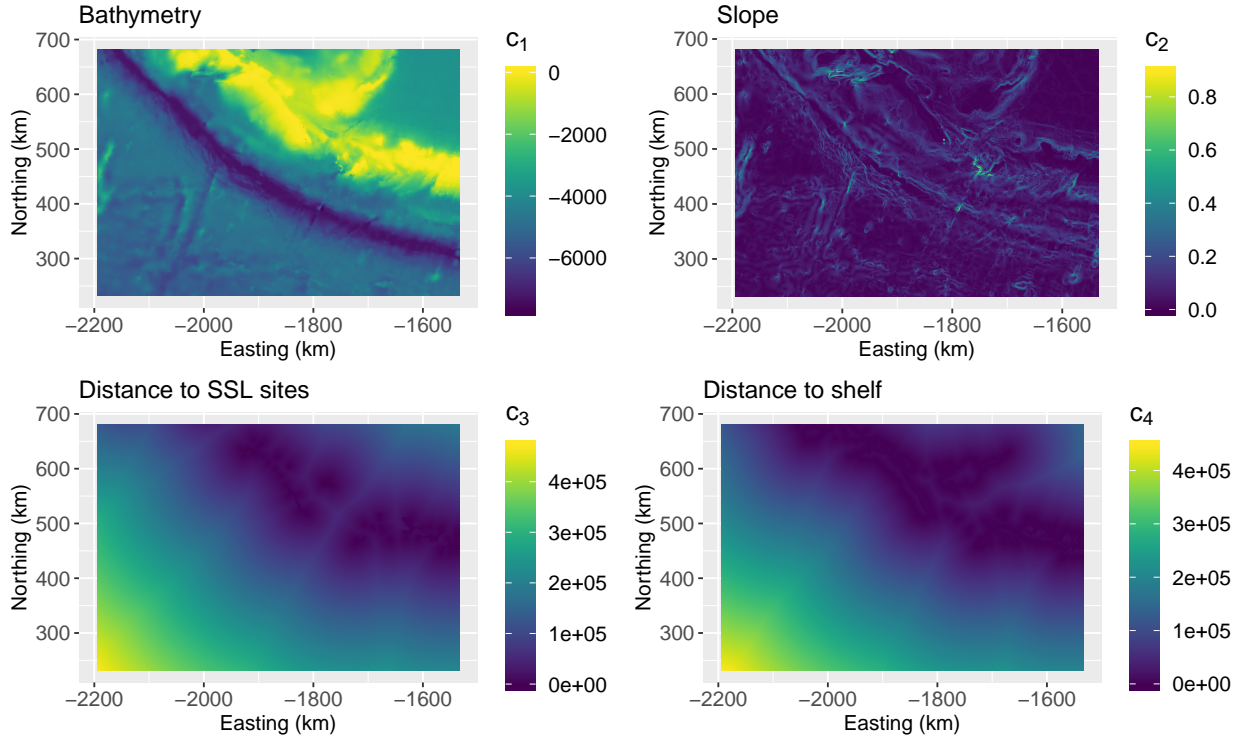


Figure 6: Covariate maps for the sea lion analysis.

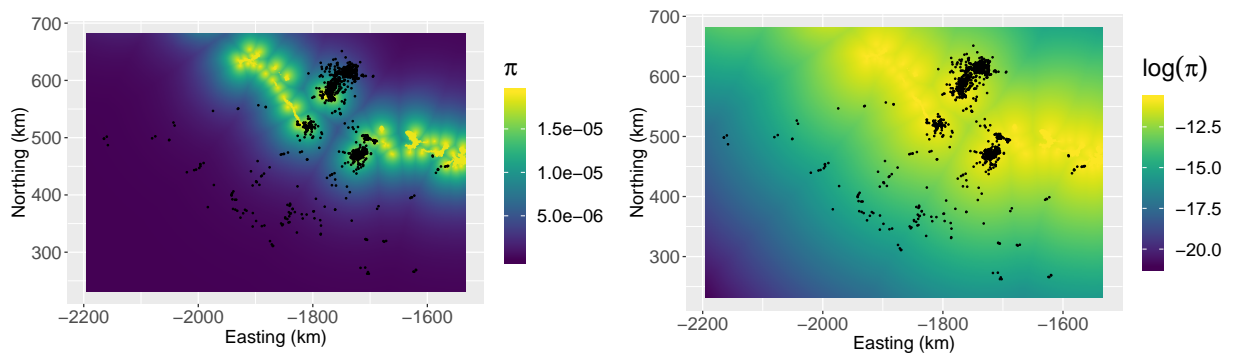
state-space model. We used the code provided by Wilson et al. (2018) to fit the CTCRW to each track, and obtained predicted locations for the times of the observations.

We then fitted the Langevin movement model to the filtered tracks, using the inference method of Section 3. Most of the computation time is needed to evaluate the gradient of each covariate at all observed locations, which took about 1.5 min on a 2GHz i5 CPU. Like in the simulation study of Section 4.2, the covariates were interpolated, so that their gradient could be evaluated at each filtered location. The point estimates and 95% confidence intervals, obtained from the equations of Section 3.2, are given in Table 1. The estimated utilisation distribution, and its logarithm (for comparison with Wilson et al., 2018), are plotted in Figure 7.

	Estimate	95% CI
$\beta_1$	$1.39 \cdot 10^{-4}$	$(-3.87 \cdot 10^{-7}, 2.79 \cdot 10^{-4})$
$\beta_2$	0.12	$(-0.14, 0.37)$
$\beta_3$	$-2.50 \cdot 10^{-5}$	$(-3.58 \cdot 10^{-5}, -1.41 \cdot 10^{-5})$
$\beta_4$	$3.47 \cdot 10^{-6}$	$(2.05 \cdot 10^{-7}, 6.73 \cdot 10^{-6})$

Table 1: Maximum likelihood estimates and 95% confidence intervals obtained with the Euler scheme for the sea lion analysis.

The 95% confidence intervals of the parameters for two of the covariates (bathymetry and slope) include zero, i.e. we cannot draw conclusions about their effects on the sea lions' movement. The effect of the distance to sites of interest,  $\beta_3$ , was estimated to be negative. This indicates that the model captures the attraction of the sea lions towards the sites of interest (rookeries and haul-out sites). On the other hand, the effect of the distance to the shelf,  $\beta_4$ , was estimated to be positive. Although this seems to indicate that the animals tend to move away from the shelf, it may also be an artifact caused by the strong collinearity between the covariates  $c_3$  and  $c_4$ .



**Figure 7:** Estimated utilisation distribution for the sea lion analysis (left), and its logarithm, for comparison with Wilson et al. (2018) (right). The black dots are the filtered sea lion locations.

We presented this analysis as a proof of concept for the Langevin movement model, and several directions could be explored further. As for any regression model, standard model selection criteria such as the AIC could be used to choose the best set of covariates. Following Wilson et al. (2018), we could also estimate the parameters separately for the three seals, to capture inter-individual heterogeneity.

## 6 Discussion

This work introduces a new model of animal movement, based on the Langevin diffusion process, that integrates the movement with space use and habitat selection. The movement model follows the idea of potential-based movement models proposed by Preisler et al. (2004) and Brillinger (2010), and it is explicitly connected to the animal’s utilisation distribution, from stationarity properties of the Langevin diffusion process (Roberts et al., 1996). If spatial covariates are available, the long-term (utilisation) distribution can be modelled with a resource selection function, embedded in the movement process, to infer habitat preferences. The Langevin movement model therefore describes animal movement in response to spatial covariates, i.e. step selection. Pseudo-likelihood methods can be used to obtain estimates of the habitat selection parameters in a classical linear model framework, from which an estimated utilisation distribution can be computed. The Langevin movement model is formulated in continuous time, and it can deal with location data collected at irregular time intervals, without the need to interpolate them. Similarly, because it models movement in continuous space (unlike the method presented by Wilson et al. 2018), the interpretation of the results is not tied to a particular space discretisation.

In this paper, we used the Euler discretization scheme to approximate the likelihood of the model. This scheme is the most widely-used method to carry out inference for discretely-observed diffusion processes, when the transition density is not analytically tractable (see Preisler et al., 2004; Brillinger, 2010; Russell et al., 2018, for applications in ecology). There exist other pseudo-likelihood approaches, and Gloaguen et al. (2018) argued that better inferences could be obtained with more refined schemes. In particular, they found that the Ozaki discretization provided more reliable results in their applications. However, the Ozaki scheme requires the evaluation of the partial derivatives of the drift, i.e. the (partial) second derivatives of  $\log \pi$  in the Langevin movement model. If  $\pi$  is modelled with a resource selection function, then this would require the evaluation of the second derivatives of the covariate fields, from Equation (5). In practice, the covariates must be interpolated to a spatially continuous function, and their second

derivatives computed using numerical methods. To compare the Euler and the Ozaki scheme, we repeated the simulation study of Section 4.1, using the Ozaki scheme for the estimation. (The results are not shown here.) The theoretical advantages of the Ozaki scheme were counterbalanced by the need of this second-order interpolation, and the Euler scheme provided more reliable estimates. Therefore, in the context of the Langevin movement model, the Euler scheme is typically more robust to numerical approximations. To avoid any discretization scheme, it would be interesting to consider exact algorithms for diffusion processes, as suggested in (Gloaguen et al., 2018). However, so far, exact methods of inference have been restricted to a small class of diffusions, and could not be readily applied to the cases presented here.

In the case study of Section 5, we used a two-stage approach to deal with the measurement error. We first fitted a state-space model, the continuous-time correlated random walk, to filter the Argos locations. Then, we fitted the Langevin movement model to the filtered tracks. There are several drawbacks to the two-stage approach. Indeed, it is difficult to propagate the uncertainty from the measurement error (although multiple imputation could be used; see e.g. Scharf et al., 2017). Besides, the two stages are not consistent, because the first stage ignores the environmental effects that are estimated in the second stage. To avoid this issue, the two steps could be integrated into a state-space model that incorporates measurement error directly on top of the Langevin movement process. The state equation of the full model is given by the transition density of the Langevin movement model, or a discretization of it (like the one given in Equation (7)). A natural choice for the observation equation would be  $\tilde{x}_i = x_i + \eta_i$ , where  $\tilde{x}_i$  is the noisy observed location,  $x_i$  is the true location, and  $\eta_i \sim N(\mathbf{0}, \sigma_{\text{obs}}^2 \mathbf{I}_2)$  models the measurement error. Under the Euler scheme, the approximate transition density is normal, and a Kalman filter can be used to compute the pseudo-likelihood of this hierarchical state-space model. This extension is conceptually straightforward, and should be done in the near future.

## Acknowledgements

TM was supported by the Centre for Advanced Biological Modelling at the University of Sheffield, funded by the Leverhulme Trust, award number DS-2014-081. We thank Paul Blackwell for discussions during the early developments of the method.

## Data accessibility

The Steller sea lion data set used in Section 5 is provided by Wilson et al. (2018).

## References

- Anderson, D. J. (1982). The home range: A new nonparametric estimation technique: Ecological archives e063-001. *Ecology*, 63(1):103–112.
- Avgar, T., Potts, J. R., Lewis, M. A., and Boyce, M. S. (2016). Integrated step selection analysis: bridging the gap between resource selection and animal movement. *Methods in Ecology and Evolution*, 7(5):619–630.
- Blackwell, P. (1997). Random diffusion models for animal movement. *Ecological Modelling*, 100(1-3):87–102.
- 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.

- Douglas Nychka, Reinhard Furrer, John Paige, and Stephan Sain (2017). *fields: Tools for spatial data*. R package version 9.6.
- 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.
- Forester, J. D., Im, H. K., and Rathouz, P. J. (2009). Accounting for animal movement in estimation of resource selection functions: sampling and data analysis. *Ecology*, 90(12):3554–3565.
- Fortin, D., Beyer, H. L., Boyce, M. S., Smith, D. W., Duchesne, T., and Mao, J. S. (2005). Wolves influence elk movements: behavior shapes a trophic cascade in yellowstone national park. *Ecology*, 86(5):1320–1330.
- Gloaguen, P., Etienne, M.-P., and Le Corff, S. (2018). Stochastic differential equation based on a multimodal potential to model movement data in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(3):599–619.
- Hanks, E. M., Hooten, M. B., Alldredge, M. W., et al. (2015). Continuous-time discrete-space models for animal movement. *The Annals of Applied Statistics*, 9(1):145–165.
- Hooten, M. B., Johnson, D. S., McClintock, B. T., and Morales, J. M. (2017). *Animal movement: statistical models for telemetry data*. CRC Press.
- 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.
- Iacus, S. M. (2009). *Simulation and inference for stochastic differential equations: with R examples*. Springer Science & Business Media.
- Johnson, D. S. and London, J. M. (2018). *crawl: an r package for fitting continuous-time correlated random walk models to animal movement data*.
- Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.
- Kessler, M., Lindner, A., and Sorensen, M. (2012). *Statistical methods for stochastic differential equations*. Chapman and Hall/CRC.
- Kranstauber, B., Kays, R., LaPoint, S. D., Wikelski, M., and Safi, K. (2012). A dynamic brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology*, 81(4):738–746.
- Long, R. A., Muir, J. D., Rachlow, J. L., and Kie, J. G. (2009). A comparison of two modeling approaches for evaluating wildlife-habitat relationships. *The Journal of Wildlife Management*, 73(2):294–302.
- Manly, B., McDonald, L., Thomas, D. L., McDonald, T. L., and Erickson, W. P. (2007). *Resource selection by animals: statistical design and analysis for field studies*. Springer Science & Business Media.
- Michelot, 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*.

- Millspaugh, J. J., Nielson, R. M., McDONALD, L., Marzluff, J. M., Gitzen, R. A., Rittenhouse, C. D., Hubbard, M. W., and Sheriff, S. L. (2006). Analysis of resource selection using utilization distributions. *The Journal of Wildlife Management*, 70(2):384–395.
- Nielson, R. M. and Sawyer, H. (2013). Estimating resource selection with count data. *Ecology and evolution*, 3(7):2233–2240.
- Potts, J. R., Bastille-Rousseau, G., Murray, D. L., Schaefer, J. A., and Lewis, M. A. (2014). Predicting local and non-local effects of resources on animal space use using a mechanistic step selection model. *Methods in ecology and evolution*, 5(3):253–262.
- Preisler, H. K., Ager, A. A., Johnson, B. K., and Kie, J. G. (2004). Modeling animal movements using stochastic differential equations. *Environmetrics*, 15(7):643–657.
- 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., Tweedie, R. L., et al. (1996). Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363.
- Russell, J. C., Hanks, E. M., Haran, M., Hughes, D., et al. (2018). A spatially varying stochastic differential equation model for animal movement. *The Annals of Applied Statistics*, 12(2):1312–1331.
- Scharf, H., Hooten, M. B., and Johnson, D. S. (2017). Imputation approaches for animal movement modeling. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):335–352.
- Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38(1/2):196–218.
- Thurfjell, H., Ciuti, S., and Boyce, M. S. (2014). Applications of step-selection functions in ecology and conservation. *Movement ecology*, 2(1):4.
- Wilson, K., Hanks, E., and Johnson, D. (2018). Estimating animal utilization densities using continuous-time markov chain models. *Methods in Ecology and Evolution*, 9(5):1232–1240.
- Worton, B. J. (1989). Kernel methods for estimating the utilization distribution in home-range studies. *Ecology*, 70(1):164–168.
- Ypma, J., Borchers, H. W., Eddelbuettel, D., and Johnson, S. (2014). R interface to nlopt.
- Zhang, Z., Sheppard, J. K., Swaisgood, R. R., Wang, G., Nie, Y., Wei, W., Zhao, N., and Wei, F. (2014). Ecological scale and seasonal heterogeneity in the spatial behaviors of giant pandas. *Integrative Zoology*, 9(1):46–60.



## A Estimators

This appendix proves that  $\hat{\beta}$  and  $\hat{\gamma}^2$  are unbiased estimators and derived the confidence intervals for the corresponding parameters.

### A.1 Unbiased estimators

We defined  $\hat{\nu}$  and  $\hat{\gamma}^2$  as

$$\begin{aligned}\hat{\nu} &= (\mathbf{D}^\top \mathbf{T}_\Delta^2 \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{T}_\Delta \mathbf{Y}, \\ \hat{\gamma}^2 &= \frac{1}{2n - J} \|\mathbf{Y} - \hat{\mathbf{Y}}\|^2.\end{aligned}$$

Following standard linear model properties,  $\hat{\nu} \sim \mathcal{N}(\gamma^2 \boldsymbol{\beta}, (\mathbf{D}^\top \mathbf{T}_\Delta^2 \mathbf{D})^{-1})$ . Besides, Cochran's theorem implies that  $\hat{\nu}$  and  $\hat{\gamma}^2$  are independent and that  $(2n - J)\hat{\gamma}^2/\gamma^2 \sim \chi^2(2n - J)$ , where  $\hat{\gamma}^2$  is an unbiased estimator of  $\gamma^2$ .

Considering  $\tilde{\beta}$  defined as  $\tilde{\beta} = \hat{\nu}/\hat{\gamma}^2$ , we have

$$\begin{aligned}\mathbb{E}[\tilde{\beta}] &= \mathbb{E}[\hat{\nu}] \mathbb{E}\left[\frac{1}{\hat{\gamma}^2}\right] \\ &= \gamma^2 \boldsymbol{\beta} \mathbb{E}\left[\frac{1}{\hat{\gamma}^2}\right] \\ &= \gamma^2 \boldsymbol{\beta} \times \frac{2n - J}{\gamma^2} \mathbb{E}\left[\frac{1}{\frac{2n - J}{\gamma^2} \hat{\gamma}^2}\right]\end{aligned}$$

From the properties given above, we have

$$\frac{1}{\frac{2n - J}{\gamma^2} \hat{\gamma}^2} \sim \text{Inv-}\chi^2(2n - J),$$

and we obtain

$$\mathbb{E}[\tilde{\beta}] = \frac{2n - J}{2n - J - 2} \boldsymbol{\beta}.$$

Thus, to obtain an unbiased estimator of  $\boldsymbol{\beta}$ , we define  $\hat{\beta} := (2n - J - 2)/(2n - J)\tilde{\beta}$ .

### A.2 Confidence intervals

As said above,  $(2n - J)\hat{\gamma}^2/\gamma^2 \sim \chi^2(2n - J)$ , so that a confidence interval for  $\gamma^2$  is given by

$$CI_\alpha(\gamma) = \left[ \hat{\gamma} \frac{2n - J}{q_{1-\alpha/2, 2n - J}}; \hat{\gamma} \frac{2n - J}{q_{\alpha/2, 2n - J}} \right],$$

where  $q_{\alpha/2, 2n - J}$  stands for the quantile of order  $\alpha/2$  from a  $\chi^2$  distribution with  $2n - J$  degrees of freedom.

It is also straightforward to derive a confidence interval or even a confidence ellipsoid using the distri-

bution of  $\tilde{\beta}$ . We first use the covariance matrix of  $\tilde{\beta}$ ,

$$\begin{aligned} \text{Cov}(\tilde{\beta}_j, \tilde{\beta}_k) &= \mathbb{E} \left[ \frac{\hat{\nu}_j \hat{\nu}_k}{\hat{\gamma}^4} \right] - \mathbb{E} \left[ \frac{\hat{\nu}_j}{\hat{\gamma}^2} \right] \mathbb{E} \left[ \frac{\hat{\nu}_k}{\hat{\gamma}^2} \right] && \text{by definition} \\ &= \mathbb{E}[\hat{\nu}_j \hat{\nu}_k] \mathbb{E} \left[ \frac{1}{\hat{\gamma}^4} \right] - \mathbb{E} \left[ \frac{1}{\hat{\gamma}^2} \right]^2 \mathbb{E}[\hat{\nu}_j] \mathbb{E}[\hat{\nu}_k] && \text{by independence of the estimators } \hat{\gamma} \text{ and } \hat{\nu}. \end{aligned}$$

We now note that

$$\begin{aligned} \mathbb{E}[\hat{\nu}_j \hat{\nu}_k] &= \text{Cov}(\hat{\nu}_j, \hat{\nu}_k) + \mathbb{E}[\hat{\nu}_j] \mathbb{E}[\hat{\nu}_k] \\ &= \gamma^2 \Upsilon_{jk} + \beta_j \beta_k \gamma^4, \end{aligned}$$

where  $\Upsilon_{jk} := [(D^\top T_\Delta^2 D)^{-1}]_{jk}$ . Moreover:

$$\begin{aligned} \mathbb{V} \left[ \frac{1}{\hat{\gamma}^2} \right] &= \frac{2(2n - J)^2}{\gamma^4 (2n - J - 2)^2 (2n - J - 4)} && \text{as } \frac{\gamma^2}{(2n - J)\hat{\gamma}^2} \sim \text{Inv-}\chi^2(2n - J) \\ \mathbb{E} \left[ \frac{1}{\hat{\gamma}^4} \right] &= \mathbb{V} \left[ \frac{1}{\hat{\gamma}^2} \right] + \mathbb{E} \left[ \frac{1}{\hat{\gamma}^2} \right]^2 = \frac{(2n - J)^2}{\gamma^4 (2n - J - 2)^2} \left( \frac{2}{2n - J - 4} + 1 \right). \end{aligned}$$

Finally, we can rewrite

$$\begin{aligned} \text{Cov}(\tilde{\beta}_j, \tilde{\beta}_k) &= \mathbb{V} \left[ \frac{1}{\hat{\gamma}^2} \right] \mathbb{E}[\hat{\nu}_j] \mathbb{E}[\hat{\nu}_k] + \mathbb{E} \left[ \frac{1}{\hat{\gamma}^4} \right] \text{Cov}(\hat{\nu}_j, \hat{\nu}_k) \\ &= \frac{(2n - J)^2}{(2n - J - 2)^2} \left\{ \frac{2\beta_j \beta_k}{2n - J - 4} + \frac{\Upsilon_{jk}}{\gamma^2} \left( 1 + \frac{2}{2n - J - 4} \right) \right\}. \end{aligned}$$

## B Simulation parameters

The simulation parameters used in Section 4.1 are given in Table A1.

	$j = 1$	$j = 2$
$\alpha_j$	6	6
$a_1^j$	0	-2
$a_2^j$	0	$\frac{\pi}{2}$
$\omega_1^j$	0.6	0.1
$\omega_2^j$	0.2	0.5
$\sigma_1^j$	0.4	0.4
$\sigma_2^j$	0.4	0.4

Table A1: Simulation parameters for Section 4.1.

## C Simulation study with irregular intervals

We repeated the second simulation scenario (Section 4.2), with irregular time intervals. Starting from the same full simulated data set ( $\Delta = 0.01$ ), we thinned the observations at random, to obtain irregularly-sampled locations. We ran two experiments: (1) mean time interval  $\bar{\Delta} = 0.05$ , (2) mean time interval

$\bar{\Delta} = 0.50$ . The estimates and 95% confidence intervals of the habitat selection parameters  $\beta_1$  and  $\beta_2$  are given in Table A2. We compare them to the results of the simulations conducted in Section 4.2 with regular intervals. The estimates obtained in the simulations with regular and irregular time intervals are very similar, and the standard errors are virtually identical.

	regular		irregular	
Mean interval	$\bar{\Delta} = 0.05$	$\bar{\Delta} = 0.5$	$\bar{\Delta} = 0.05$ (SD: 0.045)	$\bar{\Delta} = 0.5$ (SD: 0.49)
$\beta_1$	2.46 (SE: 0.68)	1.20 (SE: 0.22)	2.15 (SE: 0.69)	0.89 (SE: 0.22)
$\beta_2$	3.89 (SE: 0.65)	3.07 (SE: 0.21)	3.84 (SE: 0.65)	2.90 (SE: 0.21)

**Table A2:** Comparison of estimates and standard errors of the habitat selection parameters, for regular and irregular time intervals of simulation.

This confirms that the continuous-time formulation of the Langevin movement model can accommodate irregular time intervals, without the need to interpolate prior to the analysis.