

LA DIFFUSION DE LANGEVIN COMME MODÈLE DE MOUVEMENT EN ÉCOLOGIE POUR LE MOUVEMENT ANIMAL ET LA SÉLECTION D'HABITAT.

Pierre Gloaguen¹, Théo Michelot² et Marie-Pierre Étienne³

¹ *Agroparistech, pierre.gloaguen@agroparistech.fr*

² *University of Sheffield*

³ *Agrocampus-ouest*

Résumé. En écologie, la distribution d'utilisation décrit la probabilité relative d'utilisation d'une unité spatiale par un animal. Il est naturel de penser que cette distribution est la conséquence à long terme des décisions de mouvement à court terme de l'animal : c'est l'accumulation de petits déplacements qui, avec le temps, donne lieu à des modèles globaux d'utilisation de l'espace. Cependant, la plupart des modèles de distribution d'utilisation utilisés en écologie ignorent le mouvement sous-jacent, en supposant l'indépendance des lieux observés, ou se basent sur des règles simplistes de mouvement brownien. Nous introduisons un nouveau modèle de mouvement animal en temps continu, basé sur la diffusion de Langevin. Ce processus stochastique a une distribution stationnaire explicite, conceptuellement analogue à l'idée de distribution d'utilisation, et fournit donc un cadre intuitif pour intégrer le mouvement et l'utilisation de l'espace.

Nous modélisons la distribution stationnaire (utilisation) avec une fonction de sélection de ressources pour lier le mouvement à des covariables spatiales. Le choix de la fonction de sélection de ressource classique conduit à un schéma d'approximation naturel qui n'a besoin que des outils classiques du modèle linéaire.

Mots-clés. Diffusion de Langevin, Ecologie du mouvement, Sélection d'habitat.

Abstract. In ecology, 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 used in ecology either ignore the underlying movement, assuming the independence of observed locations, or are based on simplistic Brownian motion movement rules. 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. The choice of classical resource selection function leads to a natural approximation scheme for inference which only need the classical tools of the linear model.

Keywords. Langevin diffusion, Movement ecology, habitat selection

1 Résumé long

1.1 Contexte

Un concept crucial en écologie animale est la distribution d'utilisation, "la fonction de densité de probabilité qui donne la probabilité de trouver un animal à un endroit particulier" (Anderson, 1982). Au cours des dernières décennies, les améliorations apportées aux technologies de suivi ont produit de grandes quantités de données de localisation des animaux, à une résolution spatio-temporelle élevée. Des méthodes statistiques ont été mises au point pour estimer cette distribution d'utilisation d'après les observations de télémétrie et pour établir un lien entre les déplacements des animaux, les préférences en matière d'habitat et l'utilisation de l'espace (Hooten et al., 2017). Les approches les plus courantes pour estimer la distribution d'utilisation à partir des données de suivi sont non paramétriques et utilisent des histogrammes empiriques (Nielson and Sawyer, 2013) ou une estimation de la densité du noyau (Anderson, 1982; Worton, 1989; Fleming et al., 2015). Plus récemment, des méthodes d'interpolation faisant appel à des ponts browniens ont également été utilisées (Horne et al., 2007; Kranstauber et al., 2012). L'une des principales limites de ces méthodes réside toutefois dans le fait que l'estimation de la distribution d'utilisation est déconnectée du mouvement lui-même, car elles ignorent la structure séquentielle des données (Anderson, 1982; Worton, 1989; Nielson and Sawyer, 2013), ou se basent sur des hypothèses irréalistes sur le mouvement de l'animal (Horne et al., 2007; Kranstauber et al., 2012; Fleming et al., 2015).

Il est naturel de penser que la distribution d'utilisation est une conséquence du mouvement, qui dépend lui-même de l'environnement, de sorte que les décisions de déplacement à court terme donnent lieu à une utilisation de l'espace à long terme. Cette idée motive le développement d'approches plus mécanistes reliant le mouvement de l'animal à son environnement et, finalement, un modèle de mouvement mécaniste avec une distribution stationnaire explicite, représentant la distribution d'utilisation.

Dans ce travail, nous décrivons un nouveau modèle de mouvement mécaniste, continu dans le temps et l'espace. Nous modélisons la position de l'animal comme un processus de diffusion dont la dérive est le gradient du logarithme de sa distribution stationnaire, réunissant les idées de Brillinger (2010) et de Michelot et al. (2018). Comme dans Wilson et al. (2018), la distribution limite du processus est la distribution d'utilisation. Le modèle de mouvement que nous proposons est basé sur la diffusion de Langevin, également utilisée pour la construction d'algorithmes MCMC (Roberts and Rosenthal, 1998). L'inférence peut être réalisée à partir de données de mouvement en utilisant différentes méthodes d'estimation pour les processus de diffusion, telles que les méthodes

de pseudo-vraisemblance simples à implémenter. Nous montrons ici comment ce modèle paramétrique peut également être lié à des approches de sélection d’habitat lorsque la distribution d’utilisation est paramétrée comme une simple fonction de covariables environnementales. Les estimateurs ponctuels et les intervalles de confiance des paramètres de sélection de l’habitat peuvent facilement être dérivés dans un cadre d’inférence classique approché, celui du modèle linéaire.

1.2 Modèle de Langevin pour le mouvement

On note $\mathbf{X}_t \in \mathbb{R}^2$ la position d’un animal au temps $t \geq 0$, et $\pi : \mathbb{R}^d \rightarrow \mathbb{R}$ sa densité d’utilisation (Worton, 1989). La densité d’utilisation est la densité de probabilité telle que, pour toute zone de l’espace $A \subset \mathbb{R}^2$,

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

On modélise le processus de position $(\mathbf{X}_t)_{t \geq 0}$ grâce à la diffusion de Langevin pour la densité π . Ainsi, $(\mathbf{X}_t)_{t \geq 0}$ est solution de l’équation différentielle stochastique :

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

où \mathbf{W}_t est un mouvement Brownien standard en dimension 2, ∇ est l’opérateur gradient, γ^2 est un paramètre de vitesse et \mathbf{x}_0 est une condition initiale. Sous des conditions faibles de régularité (voir Dalalyan, 2017), l’équation (2) a une unique solution, qui est un processus de Markov en temps continu.

Ce modèle décrit ainsi le déplacement de l’animal comme la combinaison d’une dérive vers des valeurs plus élevées de sa distribution d’utilisation π (informée par le gradient de $\log \pi$) et d’une composante aléatoire donnée par le mouvement brownien. De manière cruciale, la solution est un processus de Markov à temps continu dont la distribution stationnaire asymptotique est π , telle que définie dans l’équation (1) (Roberts and Tweedie, 1996). La diffusion de Langevin est donc un choix naturel pour la base d’un modèle de mouvement animal en temps continu, dessinant une utilisation stable de l’espace de l’animal, sa distribution d’utilisation.

La Figure 1 montre deux trajectoires du processus solution de (2) une distribution d’utilisation artificielle, pour deux valeurs différentes de γ^2 . Bien que les deux trajectoires explorent l’espace à des vitesses très différentes, elles ont la même distribution stationnaires.

1.3 Inclusion de covariables de sélection d’habitat

On peut naturellement relier la distribution d’utilisation de l’animal à des covariables spatiales, en utilisant la forme paramétrique standard des fonctions de sélection de res-

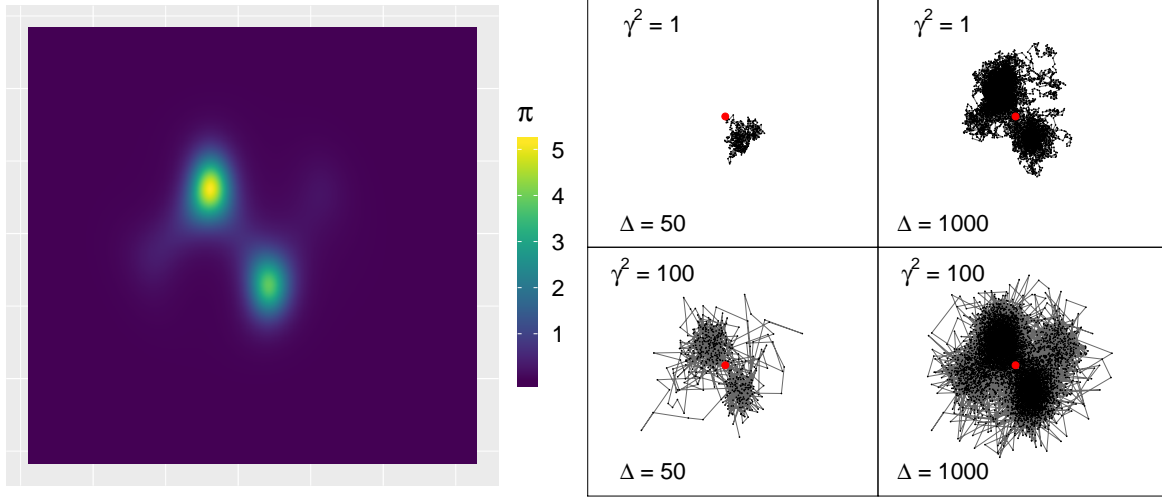


FIGURE 1 – *Gauche* : Distribution d'utilisation π . *Droite* : Trajectoires simulées à partir du modèle de mouvement Langevin de distribution stationnaire π , avec deux valeurs différentes du paramètre de vitesse γ^2 (1 et 100), après $\Delta = 50$ et $\Delta = 1000$ unités de temps. Bien que le processus avec $\gamma^2 = 1$ soit beaucoup plus lent pour explorer l'espace, les propriétés de l'équation de Langevin garantissent que les deux processus ont la même distribution stationnaire π .

sources,

$$\pi(\mathbf{x}|\boldsymbol{\beta}) = \frac{\exp\left(\sum_{j=1}^J \beta_j c_j(\mathbf{x})\right)}{\int_{\Omega} \exp\left(\sum_{j=1}^J \beta_j c_j(\mathbf{z})\right) d\mathbf{z}}, \quad (3)$$

où $c_j(\mathbf{x})$ est la valeur de la covariable j à la position \mathbf{x} , $\Omega \subset \mathbb{R}^2$ est la région d'étude, et $\boldsymbol{\beta} = (\beta_1, \dots, \beta_J)'$ est un vecteur de paramètres inconnus. La valeur de β_j mesure la force de la sélection (attraction ou évitement) pour la covariable j . Le dénominateur dans la partie droite de l'équation (3) est une constante de normalisation garantissant que $\pi(\mathbf{x}|\boldsymbol{\beta})$ est une densité de probabilité.

En injectant cette densité dans l'équation 2, on définit un modèle en temps continu de mouvement et de sélection d'habitat. Le mouvement d'un animal est ainsi modélisé comme une réponse aux caractéristiques environnementales du milieu. Ces réponses dessinant sur le long terme une utilisation stable de l'environnement, la distribution stationnaire du processus.

2 Inférence

Le processus continu des positions de l'animal, noté $(\mathbf{X}_t)_{t \geq 0}$ est observé discrètement aux temps $t_0 < t_1 < \dots < t_n$. On note ces observations $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n)$. On considère J

covariables spatiales c_1, \dots, c_J , mesurées sur une grille de l'espace. $\boldsymbol{\theta}$ désigne le vecteur de tous les paramètres du modèle de mouvement de Langevin défini dans la section 1.2, c'est-à-dire $\boldsymbol{\theta} = (\beta_1, \dots, \beta_J, \gamma^2)$. Cette section décrit une méthode d'inférence permettant d'estimer $\boldsymbol{\theta}$ à partir de données de télémétrie et d'habitats.

2.1 Approximation de la vraisemblance

La vraisemblance de $\boldsymbol{\theta}$, peut être exprimée à l'aide de la *densité de transition* du processus $(\mathbf{X}_t)_{t \geq 0}$. La densité de transition est la fonction de densité de probabilité de la variable aléatoire $\mathbf{X}_{t+\Delta}$ étant donné que $\mathbf{X}_t = \mathbf{x}_t$, notée $q_\Delta(\mathbf{x}|\mathbf{x}_t, \boldsymbol{\theta})$. Par la propriété de Markov de la solution de l'équation (2), et en supposant que la première position est déterministe, la fonction de vraisemblance est

$$L(\boldsymbol{\theta}; \mathbf{x}_{0:n}) = \prod_{i=0}^{n-1} q_{\Delta_i}(\mathbf{x}_{i+1}|\mathbf{x}_i, \boldsymbol{\theta}), \quad (4)$$

où $\mathbf{x}_{0:n}$ désigne l'ensemble des observations, et $\Delta_i = t_{i+1} - t_i$.

Comme discuté dans Gloaguen et al. (2018), dans de nombreux cas pratiques, il n'existe pas d'expression analytique pour la densité q_Δ , ainsi la vraisemblance $L(\boldsymbol{\theta}; \mathbf{x}_{0:n})$ ne peut pas être évaluée. Pour contourner ce problème, les approches de pseudo-vraisemblance peuvent être utilisées comme approximations.

L'approche la plus classique est l'approche par schéma d'Euler (Iacus, 2009). Dans cette discrétisation, la densité de transition de la diffusion de Langevin est approchée, entre t_i et t_{i+1} , pour $i = 0, \dots, n-1$, par la densité de la loi Gaussienne suivante :

$$\mathbf{X}_{i+1}|\{\mathbf{X}_i = \mathbf{x}_i\} = \mathbf{x}_i + \frac{\gamma^2 \Delta_i}{2} \nabla \log \pi(\mathbf{x}_i|\boldsymbol{\beta}) + \varepsilon_{i+1}, \quad \varepsilon_{i+1} \stackrel{ind}{\sim} N(\mathbf{0}, \gamma^2 \Delta_i \mathbf{I}_d), \quad (5)$$

où \mathbf{I}_2 est la matrice identité.

Cette écriture permet ainsi d'écrire une pseudo vraisemblance pour tout paramètre $\boldsymbol{\theta}$. La qualité d'approximation de cette pseudo vraisemblance sera bonne quand Δ_i est petit.

2.2 Estimateurs

La pseudo vraisemblance issue de la discrétisation d'Euler peut être maximisée numériquement pour obtenir un estimateur de $\boldsymbol{\theta}$. Cependant, si π est donnée par l'équation (3), le modèle de mouvement discrétisé peut être écrit en termes de modèle linéaire et l'estimation $\hat{\boldsymbol{\theta}}$ peut simplement être obtenu en utilisant des équations de modèle linéaires standard. Cet estimateur permet d'obtenir des intervalles de confiance sur les estimateurs et de statuer sur l'influence des variables environnementales sur les mouvements de l'individu.

3 Application au lion de mer de Steller

On ajuste le modèle de mouvement de Langevin à un jeu de données décrit par Wilson et al. (2018), collecté sur des lions de mer de Steller (*Eumetopias jubatus*) en Alaska. L'ensemble de données comprend trois trajectoires, obtenues à partir de trois individus différents, pour un total de 2672 positions Argos. En plus des positions, Wilson et al. (2018) fournit quatre covariables spatiales, à une résolution de 1 km : bathymétrie (c_1), pente (c_2), distance à la colonie (c_3) et distance au plateau continental (c_4).

L'estimation faite ici permet de conclure que le mouvement est structuré par la bathymétrie, la distance à la colonie, ainsi qu'à la distance au plateau continental. L'estimation de la densité d'utilisation est montrée sur la Figure 2

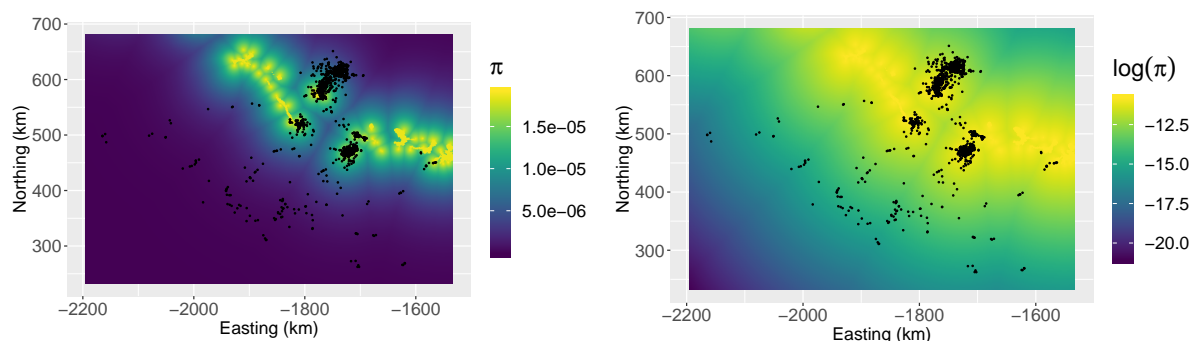


FIGURE 2 – Estimation de la densité d'utilisation des lions de mers (ainsi que de son logarithme). Les points noirs représentent les points GPS.

Références

- Anderson, D. J. (1982). The home range : A new nonparametric estimation technique : Ecological archives e063-001. *Ecology*, 63(1) :103–112.
- 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 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.
- 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.
- 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.
- Michelot, T., Blackwell, P. G., and Matthiopoulos, J. (2018). Linking resource selection and step selection models for habitat preferences in animals. *Ecology*. DOI : 10.1002/ecy.2452.
- Nielson, R. M. and Sawyer, H. (2013). Estimating resource selection with count data. *Ecology and evolution*, 3(7) :2233–2240.
- 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*, 2(4) :341–363.
- 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.