



LABORATOIRE
JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE



UNIVERSITY OF
COPENHAGEN

CNRS-LJK – University of Copenhagen PhD subject

CNRS Advisor: Adeline Samson (**AS**), Laboratoire Jean Kuntzmann UMR 5224 CNRS, Université Grenoble Alpes.

Email: adeline.leclercq-samson@univ-grenoble-alpes.fr

UPCH Advisor: Susanne Ditlevsen (**SD**), Department of Mathematical Sciences.

Email: susanne@math.ku.dk

UPCH Co-advisor: Mads Peter Heide-Jørgensen (**MPHJ**), Greenland Institute of Natural Resources. Email: mhj@ghsdk.dk

Context of the scientific project

Statistical analysis of whales movement trajectories

Movement ecology under climate changes: a very important scientific problem! It is both exciting and daunting to understand and predict the ongoing ecosystem changes in the Arctic. Exciting because huge data sets are now collected by modern empirical methods, including detailed information on behavior of marine mammals together with extensive measurements on local characteristics. Information we could only dream of just a few years back. Daunting because over the past 30 years sea ice and climate has changed rapidly in the Arctic generating unpredictable effects that are cascading through the marine food chains even affecting the top predators.

Explosion of new data in marine ecology. Modern methods for data sampling from instrumented Arctic marine mammals contribute unprecedented details on animal behavior in the wild.

Challenge of developing new statistical methods. These new datasets are challenging and require development of new statistical and data science methods to unravel biological mechanisms. Mathematical modeling and statistical challenges include:

- dependence structures in the data.
- different sources of variation: within and between individuals, over the life span of individuals, yearly variations, trends over time due to climate changes, temporal and spatial variations.
- jointly modeling of different data types, e.g., sound production modelled with stochastic point processes, behavioral data that are continuous stochastic processes.
- estimation of latent biological variables (e.g., stress levels) from multidimensional behavioral data.

PhD subject: Movement trajectories.

Effect of the changing Artic environment on movement trajectories. The central question in this WP is to quantify and estimate the effect of the changing Arctic environment on marine mammal behaviors, especially on their annual migration patterns. As suggested by Dinsdale et al 2019, we will quantify if animals adjust their foraging due to the warming of ocean temperatures. Can we estimate if they follow prey towards colder water? The data that we will use consist of location coordinates of bowhead whales and narwhals obtained with telemetry data and environmental factors such as sea surface temperature measurements (see Fig 1).

Stochastic models along the migration route. We will develop statistical models relating movement processes of animals along their migration route and spatial habitat conditions and preferences, as proposed by Hooten et al 2010, Michelot et al 2019, Johnson et al 2021, to quantify the effect of the changing Arctic environment on animal behaviors.

Stochastic diffusion on the animal location. Stochastic models provide a natural framework to analyze animal movements. Stochastic models include continuous time Markov Chains on a discrete space (Hooten et al 2010, Hanks et al 2015) or stochastic diffusions on a continuous space (Preisler et al 2013, Gloaguen et al 2018, Michelot et al 2019). In this project, we focus on stochastic diffusion models, expressed either directly on the position $X(t)$ or the velocity $V(t)$ at time t . A stochastic differential equation (SDE) of the position $X(t)$ can be defined as:

$$dX(t) = \mu(X(t), S, \theta)dt + \sigma(X(t), t)dW(t)$$

where the function $\mu(x, S, \theta)$ models the direction preference depending on the position x but also spatial covariates denoted $S = (S_1, \dots, S_K)$ (such as sea temperature, distance to coast, salinity, ice coverage, presence of preys), and unknown parameters θ . The second term models the stochasticity in the process, including both environmental and individual randomness. The drift should include a reflection term to constrain the animal in space, as the sea domain is bounded, leading to reflected SDE (Hanks et al 2017).

Effect of environmental spatial maps. Effects of environmental covariates can be quantified through differential potential functions H of the maps, modeling different behaviors (foraging, migration, etc), through a mixture drift:

$$\mu(x, S, \theta) = \sum_{k=1}^K \gamma_k H'(S_k(x), \theta)$$

where γ_k are weights of the different behaviors. Then, time-varying coefficients $\theta(t)$ quantifies the effect of the changing environment, as recently proposed by Michelot et al 2021. Finally, when the drift is applied to the velocity, an additional deterministic equation modeling the position is added leading to a hypoelliptic SDE (Hanks et al 2017, Ditlevsen and Samson, 2019). Hypoelliptic means that the noise is degenerate, which complicates the statistical analysis.

Statistical challenges of movement diffusion processes. These models generalize the class of reflected Langevin diffusions, with a mixture of potential functions as the drift, which are possibly hypoelliptic. Estimation methods are difficult, the solution of the SDE being non explicit and the transition densities are unknown, so the likelihood function is not available. We have recently developed new estimation methods for complex SDEs, combining discrete numerical schemes (1.5 scheme, Splitting scheme), and computational algorithms (SAEM, MCMC, particle filtering). However, including simultaneously spatial covariates, mixture of potentials and a reflected drift is an additional computational challenge, where we need to augment the constrained process in a Markov Chain Monte Carlo procedure.

Milestones of the PhD are the followings:

Milestone 1: Provide an estimation algorithm to estimate parameters of an elliptic diffusion with a mixture potential as a drift. Prove the theoretical properties of this algorithm.

Milestone 2: Provide an estimation algorithm for hypoelliptic diffusion with reflected drift, with theoretical properties.

Milestone 3: Apply both algorithms to spatio-temporal trajectories of the animals.

Milestone 4: Relate the estimated environmental map in WP1 to movement data (together with PhD 1).

Expected results. The quantification of changes in movement trajectories is of direct importance for whale biology but also for environmental and climate science. Our central hypothesis is that changes in the environment have a direct effect on marine mammal behaviors and on their survival.

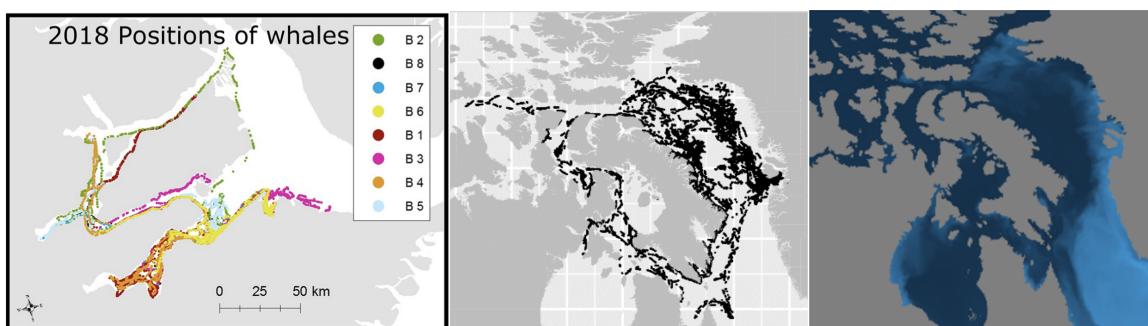


Fig 1. Left: Trajectories of narwhals in Scoresbysund in East Greenland. The colors indicate different individuals. Middle: Migration of bowhead whales between West Greenland and Canada. Right: Temperature map for the same area as in the middle.

Joint supervision

Another PhD student will be based in Copenhagen, working on ecological archive in narwhal tusks. We will have weekly online meetings and meet in person all four (PIs and PhD students) twice a year, once in Copenhagen and once in Grenoble. Furthermore, the UCPH co-investigator MPHJ will supervise the biological parts of the project. He will participate in the online meetings as needed, and join the physical meetings in Copenhagen. The students will be physically together in either Grenoble or Copenhagen for 4 months each year.

Timeline of travels:

October 2022: PhD enrollment. PhD1 will be enrolled in Copenhagen, PhD2 will be enrolled in Grenoble. Start-up meeting in Copenhagen with all present, 1 week.

February-March 2023/2024/2025: PhD1 will visit Grenoble for 2 months.

February 2023/2024/2025: Physical meeting in Grenoble, PIs and PhDs, one week.

September-October 2023/2024: PhD2 will visit Copenhagen for 2 months.

October 2023/2024: Physical meeting in Copenhagen, PIs, PhDs, co-investigator, 1 week.

May-June 2025: PhD2 will visit Copenhagen for 2 months.

October-December 2025: Evaluation and defenses of PhD theses.

References

Dinsdale and Salibian-Barrera. Modelling ocean temperatures from bio-probes under preferential sampling, *A. App. Statistics*, 2019.

Ditlevsen and Samson: Estimation in the partially observed stochastic Morris-Lecar neuronal model with particle filter and stochastic approximation methods. *Annals of Applied Statistics*, 2014, 8(2): 674-702.

Ditlevsen and Samson: Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations, *JRSS B*, 2019, 81, 361-384.

Gloaguen, Etienne, Le Corff: Stochastic differential equation based on a multimodal potential to model movement data in ecology. *JRSS C*, 2018, 67, 599-619.

Hanks, Hooten, Alldredge: Continuous-time discrete space models for animal movement, *A. App Statistics*, 2015, 9, 145-165.

Hanks, Johnson, Hooten: Reflected stochastic differential equation models for constrained animal movement, *J Agri Biological Environ Stat*, 2017, 22, 353-372.

Heide-Jørgensen, Blackwell, Tervo, Samson, Garde, Hansen, Ngo, Conrad, Trinhammer, Schmidt, Sinding, Williams, Ditlevsen: Behavioral response study on seismic airgun and vessel exposures in narwhals. *Frontiers in Marine Science*, 2021, 8:658173. doi: 10.3389/fmars.2021.658173.

Heide-Jørgensen, Blackwell, Williams, Sinding, Skovrind, Tervo, Garde, Hansen, Nielsen, Ngo, Ditlevsen: Some like it cold: Temperature dependent habitat selection by narwhals. *Ecology and Evolution*, 2020, 10(15), 8073-8090. <https://doi.org/10.1002/ece3.6464>.

Hooten, Johnson, Hanks, Lowry, Agent-based inference for animal movement and selection. *J. Agri. Biol. Environ.Stat.* 2010, 15, 523-538.

Johnson, Pelland, Sterling: A continuous-time semi-Markov model for animal movement in a dynamic environment, *A. App. Statistics*, 2021, 15, 797-812.

Michelot, Glennie, Harris, Thomas: Varying-coefficient stochastic differential equations with applications in ecology, *J. Agric Biol Envir Statistics*, 2021, 26, 446-463.

Michelot, Gloaguen, Blackwell, Etienne: The Langevin diffusion as a continuous-time model of animal movement and habitat selection. *Meth Eco Evolution*, 2019, 10, 1894-1907.

Michelot, Gloaguen, Etienne: Reproducing the analysis presented in the Langevin diffusion as a continuous-time model of animal movement and habitat selection. Zenodo, 2019.

Preisler, Ager, Wisdom: Analyzing animal movement patterns using potential functions, Ecosphere, 2013, 4, 1-13.

Tervo, Blackwell, Ditlevsen, Conrad, Samson, Garde and Heide-Jørgensen: Narwhals react to ship noise and airgun pulses embedded in background noise. Biology Letters, 2021, 17(11), <https://doi.org/10.1098/rsbl.2021.0220>

Analyse statistique des trajectoires de baleines

L'écologie du mouvement sous les changements climatiques est un problème scientifique très important. Il est à la fois passionnant et décourageant de comprendre et de prévoir les changements en cours dans l'écosystème de l'Arctique. Excitant parce que d'énormes ensembles de données sont maintenant collectés par des méthodes empiriques modernes, y compris des informations détaillées sur le comportement des mammifères marins ainsi que des mesures extensives sur les caractéristiques locales

Les méthodes modernes d'échantillonnage des données provenant de mammifères marins équipés d'instruments fournissent des détails sans précédent sur le comportement des animaux dans la nature.

Ces nouveaux ensembles de données constituent un défi et nécessitent le développement de nouvelles méthodes statistiques et de science des données pour démêler les mécanismes biologiques. La modélisation mathématique et les défis statistiques comprennent

- les structures de dépendance dans les données.
- les différentes sources de variation : au sein des individus et entre eux, au cours de la vie des individus, les variations annuelles, les tendances dans le temps dues aux changements climatiques, les variations temporelles et spatiales.
- la modélisation conjointe de différents types de données, par exemple, la production de sons modélisée par des processus ponctuels stochastiques, les données comportementales qui sont des processus continus,
- estimation de variables biologiques latentes (par exemple, les niveaux de stress) à partir de données comportementales multidimensionnelles.

Sujet de thèse : Trajectoires de mouvement.

La question centrale de cette thèse est de quantifier et d'estimer l'effet du changement de l'environnement arctique sur les comportements des mammifères marins, en particulier sur leurs schémas de migration annuelle. Comme suggéré par Dinsdale et al 2019, nous allons quantifier si les animaux ajustent leur recherche de nourriture en raison du réchauffement des températures océaniques. Pouvons-nous estimer s'ils suivent leurs proies vers des eaux plus froides ?

Les données que nous utiliserons consistent en des coordonnées de localisation des baleines boréales et des narvals obtenues grâce à des données de télémétrie et des facteurs environnementaux tels que des mesures de la température de surface de la mer (voir Fig 1).

Nous développerons des modèles statistiques mettant en relation les processus de déplacement des animaux le long de leur route de migration et les conditions et préférences spatiales en matière d'habitat, comme proposé par Hooten et al 2010, Michelot et al 2019, Johnson et al 2021, afin de quantifier l'effet de l'évolution de l'environnement arctique sur les comportements animaux.

Les modèles stochastiques fournissent un cadre naturel pour analyser les mouvements des animaux. Les modèles stochastiques comprennent les chaînes de Markov à temps continu sur un espace discret (Hooten et al 2010, Hanks et al 2015) ou les diffusions stochastiques sur un espace continu (Preisler et al 2013, Gloaguen et al 2018, Michelot et al 2019). Dans ce projet, nous nous concentrerons sur les modèles de diffusion stochastique, exprimés soit directement

sur la position $X(t)$ ou la vitesse $V(t)$ au temps t . Une équation différentielle stochastique (EDS) de la position $X(t)$ peut être définie comme :

$$dX(t) = \mu(X(t), S, \theta) dt + \sigma(X(t), t) dW(t)$$

où la fonction $\mu(x, S, \theta)$ modélise la préférence de direction en fonction de la position x mais aussi des covariables spatiales notées $[S=(S_1, \dots, S_K)]$ (telles que la température de la mer, la distance à la côte, la salinité, la couverture de glace, la présence de proies), et des paramètres inconnus θ . Le second terme modélise la stochasticité du processus, incluant le hasard environnemental et individuel. La dérive doit inclure un terme de réflexion pour contraindre l'animal dans l'espace, car le domaine marin est borné, ce qui conduit à une EDD réfléchie (Hanks et al 2017).

Les effets des covariables environnementales peuvent être quantifiés par des fonctions potentielles différentielles H des cartes, modélisant différents comportements (recherche de nourriture, migration, etc.), à travers une dérive de mélange :

$$\mu(x, S, \theta) = \sum_{k=1}^K [\gamma_k H(S_k(x), \theta)]$$

où γ_k sont les poids des différents comportements. Ensuite, les coefficients variables dans le temps $\theta(t)$ quantifient l'effet de l'environnement changeant, comme récemment proposé par Michelot et al 2021. Enfin, lorsque la dérive est appliquée à la vitesse, une équation déterministe supplémentaire modélisant la position est ajoutée, conduisant à un SDE hypoelliptique (Hanks et al 2017, Ditlevsen et Samson, 2019). Hypoelliptique signifie que le bruit est dégénéré, ce qui complique l'analyse statistique.

Ces modèles généralisent la classe des diffusions de Langevin réfléchies, avec un mélange de fonctions potentielles comme dérive, qui sont éventuellement hypoelliptiques. Les méthodes d'estimation sont difficiles, la solution de l'EDS étant non explicite et les densités de transition étant inconnues, la fonction de vraisemblance n'est pas disponible. Nous avons récemment développé de nouvelles méthodes d'estimation pour les EDS complexes, combinant des schémas numériques discrets (schéma 1.5, schéma de fractionnement), et des algorithmes de calcul (SAEM, MCMC, filtrage particulaire). Cependant, l'inclusion simultanée de covariables spatiales, d'un mélange de potentiels et d'une dérive réfléchie constitue un défi informatique supplémentaire, où nous devons augmenter le processus contraint dans une procédure de Monte Carlo par chaîne de Markov.

Le doctorant sera basé au LJK à Grenoble. Des échanges auront lieu avec un autre doctorant basé à Copenhague. Le doctorant sera basé dans l'Unité CNRS 5224 Laboratoire Jean Kuntzmann à Grenoble. Ce financement CNRS vise à soutenir les échanges avec l'Université de Copenhague au Danemark. Un deuxième doctorant sera basé à Copenhague et analysera d'autres données mesurées sur les baleines narvalles. Le financement CNRS permettra de financer des séjours longs des deux doctorants, chaque année, dans la ville jumelle (Grenoble/Copenhague)..