Offre de stage Master 2

Quantification et calibration des incertitudes dans les modèles spatiaux de distribution des cétacés

Organisme d'accueil : La Rochelle Université

Domaine : Mathématiques appliquées, statistiques, modélisation — écologie marine

Durée : 6 mois

Période : février–septembre 2027 (adaptable)

Lieu : CMAP polytechnique (Palaiseau) ou Banulsdesign (Larmor-Baden), selon le profil et les disponibilités

Gratification : au taux légal en vigueur, revalorisé chaque année en janvier

Encadrement : Auriane Virgili (Share The Ocean), Olivier Le Maître (CMAP polytechnique), Renaud Bañuls (Banulsdesign), Matthieu Authier (La Rochelle Université)

Candidature : CV + lettre de motivation à auriane@sharetheocean.earth, renaud@sharetheocean.earth, olivier.le-maitre@polytechnique.edu, matthieu.authier@univ-lr.fr

---

Contexte scientifique

La cartographie de la distribution des cétacés repose sur des modèles de distribution des espèces (Species Distribution Models, SDM), ajustés notamment à partir de campagnes d'observation visuelle en mer. Ces prédictions alimentent directement des applications de gestion et de conservation, telles que l'évaluation du risque de collision avec les navires. Leur utilité dépend donc autant de leur qualité prédictive que de la fiabilité des incertitudes qui les accompagnent.

Cette question est particulièrement délicate dans le cas des campagnes visuelles, où les observations résultent de deux mécanismes distincts : la distribution effective des animaux et leur probabilité de détection, qui dépend notamment des conditions d'observation et du comportement des cétacés. Une faible fréquence d'observation peut ainsi résulter d'une faible densité, d'une détection imparfaite ou d'une combinaison des deux. La séparation de ces effets soulève un problème d'identifiabilité statistique, qui conditionne la fiabilité des estimations de densité et de leurs incertitudes.

Les modèles actuellement utilisés reposent sur des modèles additifs généralisés (GAM), implémentés avec le package R mgcv. Bien que ce cadre permette de quantifier certaines composantes de l'incertitude, leur représentation et leur propagation dans les

prédictions spatiales méritent d'être approfondies, notamment en présence de dépendances spatiales résiduelles.

Le stage vise à analyser les incertitudes générées par ces modèles, à étudier des formulations alternatives issues de la statistique spatiale et à comparer leur capacité à fournir des estimations fiables et correctement calibrées.

Objectifs et missions

1. Analyse et diagnostic des incertitudes des modèles existants

Le premier objectif consiste à examiner la formulation statistique des SDM actuels et à identifier les principales sources d'incertitude : la variabilité des observations, l'estimation des paramètres, la détection imparfaite et la prédiction dans les régions peu échantillonnées.

Une attention particulière sera portée à la distinction entre les intervalles de confiance sur les densités estimées, qui caractérisent l'incertitude liée à l'inférence sur la distribution spatiale sous-jacente, et les intervalles prédictifs, qui décrivent la variabilité des observations futures en intégrant l'incertitude d'estimation et la variabilité intrinsèque du processus d'observation.

Cette analyse permettra de préciser les composantes d'incertitude effectivement représentées par les GAM, les hypothèses sous-jacentes et les limites éventuelles des méthodes disponibles dans mgcv.

2. Étude et mise en œuvre d'un modèle statistique spatial alternatif

Le deuxième objectif vise à étudier un cadre permettant de représenter explicitement les dépendances spatiales et les incertitudes associées. Une approche bayésienne spatiale, reposant par exemple sur un champ aléatoire gaussien modélisé par une SPDE et sur une inférence approchée de type INLA, pourra être envisagée.

Le travail consistera à formuler et implémenter un modèle adapté aux données disponibles, en distinguant l'incertitude sur le champ spatial latent décrivant la distribution des cétacés et celle affectant les observations. Selon l'avancement du stage, la sensibilité aux hypothèses de covariance spatiale et aux choix de régularisation pourra également être étudiée.

3. Évaluation comparative et calibration des incertitudes

Les approches GAM et spatiale alternative seront comparées sur un jeu de données commun, en considérant à la fois la qualité des estimations ponctuelles et celle des incertitudes associées.

L'évaluation portera notamment sur la couverture empirique, la largeur des intervalles et des scores probabilistes adaptés. La calibration des intervalles prédictifs pourra être examinée à partir des observations disponibles, tandis que celle des intervalles de confiance sur les densités, non directement observables, pourra être étudiée à partir d'expériences menées sur des données simulées.

En prolongement, selon l'avancement du stage, des procédures de validation tenant compte des dépendances spatiales, notamment par blocs spatiaux ou par campagnes indépendantes, pourront être explorées.

Résultats attendus

Le stage devra aboutir à une analyse critique des incertitudes des SDM actuels, à la mise en œuvre d'une formulation statistique alternative et à une comparaison documentée de leurs propriétés prédictives et de leur calibration.

Les résultats permettront de formuler des recommandations méthodologiques pour la production de cartes de densité des cétacés accompagnées d'indicateurs d'incertitude fiables, en vue de leur utilisation dans les modèles d'évaluation du risque de collision.

Profil recherché

- Étudiant·e en Master 2 (ou dernière année d'école d'ingénieur) en mathématiques appliquées ou statistiques.
- Bases solides en statistiques et en modélisation ; un intérêt ou une première expérience en statistiques spatiales ou en inférence bayésienne est un plus.
- Programmation en R et/ou Python.
- Appétence pour la recherche, autonomie et rigueur.
- Goût pour l'interdisciplinarité et pour l'application des mathématiques à la conservation ; aucune connaissance préalable du milieu marin n'est requise.

Compétences développées

- Quantification, décomposition et calibration des incertitudes de prédiction.
- Statistique spatiale appliquée à un problème réel de conservation.
- Comparaison rigoureuse de cadres de modélisation à partir de données réelles issues de campagnes en mer.
- Interface entre statistiques, mathématiques appliquées et écologie ; communication de résultats scientifiques.

Candidature

Envoyer un CV et une lettre de motivation à auriane@sharetheocean.earth, renaud@sharetheocean.earth, olivier.le-maitre@polytechnique.edu, matthieu.authier@univ-lr.fr. Les candidatures sont examinées au fil de l'eau, jusqu'à ce que le stage soit pourvu. Des perspectives de poursuite pourront être évoquées lors de l'entretien.