

# **Rapport d'activité**

**Mars 2023**

Pierre Gloaguen, section CNECA n°3

# Table des matières

<b>1</b>	<b>Présentation du candidat</b>	<b>4</b>
1.1	État civil . . . . .	4
1.2	Cursus professionnel, fonctions exercées, mobilité . . . . .	4
1.3	Diplômes et formations complémentaires . . . . .	4
	Diplômes . . . . .	4
<b>2</b>	<b>Activités d'enseignement</b>	<b>6</b>
2.1	Participation à l'enseignement de l'établissement . . . . .	6
2.1.1	Services d'enseignement, réalisé pédagogique . . . . .	6
2.1.2	Démarches pédagogiques, responsabilités assumées . . . . .	9
2.1.3	Création - Innovation - Evolution - Actualisation – Prospective . . . . .	9
2.1.4	Encadrement d'étudiants pour l'enseignement (hors activités de recherche) . . . . .	9
2.2	Participation à l'enseignement en France hors de l'établissement . . . . .	10
2.3	Missions d'enseignement à l'étranger . . . . .	10
2.4	Création de documents pédagogiques et publications d'enseignement . . . . .	10
<b>3</b>	<b>Activités de recherche et de développement</b>	<b>12</b>
3.1	Orientation générale . . . . .	12
	Contexte . . . . .	12
	Inférence dans des modèles de Markov caché généraux: définition d'un nouvel estimateur avec garanties théoriques . . . . .	12
	Problématiques appliquées . . . . .	14
	Animation de recherche . . . . .	15
3.2	Cadre structurel de la recherche et du développement . . . . .	16
3.3	Objectifs - Thème(s) (genèse, état actuel, perspective) - Insertion dans les programmes locaux, régionaux, nationaux, européens et internationaux . . . . .	16
3.4	Encadrement de la recherche dans le cadre de l'activité de recherche per- sonnelle . . . . .	17
3.5	Développement rural, agro-industriel . . . . .	18
3.6	Recherche dans le cadre de la coopération internationale . . . . .	18
3.7	Publications scientifiques et valorisation . . . . .	18

<b>4</b>	<b>Autres activités</b>	<b>19</b>
4.1	Activités de prestation de service et d'expertise . . . . .	19
4.2	Responsabilité, fonctions et réseaux au service de l'établissement (y compris à l'international) . . . . .	19
4.3	Participation à des jurys (membre ou rapporteur) internes et externes . .	19
4.4	Responsabilités et fonctions externes à l'établissement . . . . .	19
4.5	Rayonnement national et international . . . . .	20
4.5.1	Présentations en conférences et ateliers de recherche . . . . .	20
4.5.2	Présentations en séminaire de recherche . . . . .	20
4.6	Appartenance à des commissions et des groupes de travail institutionnels	20
<b>5</b>	<b>Réflexion sur les activités et perspectives</b>	<b>21</b>

# 1 Présentation du candidat

## 1.1 État civil

- Pierre GLOAGUEN, né le 18 Septembre 1987 à Vannes (56);
- Maître de conférences du ministère de l'enseignement supérieur agricole ( depuis Septembre 2018;
- Échelon 5;
- Section CNECA n°3;
- Etablissement: Agroparistech;
- Adresse: 22 place de l'Agronomie, 91400, Palaiseau;
- Courriel: pierre.gloaguen@agroparistech.fr;
- Téléphone: 01 89 10 09 49;

## 1.2 Cursus professionnel, fonctions exercées, mobilité

- *09/2018- ...*: Maître de conférences en Statistiques à Agroparistech;
- *06/2017 - 08/2018*: Post Doctorat à l'Université Bretagne Sud.
- *03/2016 - 03/2017*: Post Doctorat à l'Institut National de la Recherche Agronomique (INRA), AgroParisTech.
- *11/2012 - 11/2015*: Thèse à l'Institut Français de Recherche pour l'Exploitation de la Mer (IFREMER), Nantes. **Modélisation mécaniste et stochastique des trajectoires pour l'halieutique.**

## 1.3 Diplômes et formations complémentaires

### Diplômes

- **Doctorat** obtenu en Décembre 2015, décerné par l'Agrocampus-Ouest, spécialité Écologie. Thèse encadrée par Stéphanie Mahévas, Etienne Rivot et Marie-Pierre

Etienne. Jury de thèse composé de

- *Rapporteurs*: Olivier Gimenez (Directeur de Recherche au Centre d'Ecologie Fonctionnelle et Évolutive); et Adeline Leclercq Samson (Professeure en Statistiques à l'Université Grenoble Alpes);
- *Examineurs*: Didier Gascuel (Professeur Agrocampus-Ouest), Juan Manuel Morales (Enseignant chercheur , Universidad Nacional del Comahue) et Samuel Soubeyrand (Directeur de Recherche INRAe);
- **Master** en Mathématiques, Statistiques et applications, obtenu en Juin 2012, décerné par l'Université Montpellier II (en partenariat avec SupAgro Montpellier);
- **Licence** en Mathématiques, Statistiques et applications, obtenu en Juin 2008, décerné par l'Université Rennes I;

## 2 Activités d'enseignement

### 2.1 Participation à l'enseignement de l'établissement

#### 2.1.1 Services d'enseignement, réalisé pédagogique

- Enseignant dans l'UFR de mathématiques du département MMIP;
- Plus ou moins 192 heures par an réparties entre:
  - Les 3 années du cursus ingénieur Agroparistech;
  - Le Master Biologie, écologie et évolution de l'Université Paris-Saclay, cohabilité par Agroparistech;
  - Le Master d'écologie Tropicale cohabilité Agroparistech et Université Antilles/Guyane
- Ci dessous, le descriptif de mes principaux enseignements des 4 dernières années. Le décompte exact est fourni en annexe B. Les remplacements ponctuels ne sont pas recensés.

#### Interventions dans le cursus ingénieur 1A

- **Modélisation mathématique et ses applications : statistiques;**
  - Cursus: Tronc commun;
  - Groupes de 30 étudiants;
  - Chargé de TD;
  - Outils de base de statistique inférentielle pour l'ingénieur;
  - Évaluation en contrôle continu et examen final écrit;
  - Collaborateurs: Responsable du cours et 5 autres chargés de TD;

## **Interventions dans le cursus ingénieur 2A**

- **Sciences des données : apprentissage statistique;**
  - Cursus: Tronc commun;
  - Groupes de 30 étudiants;
  - Chargé de cours (avant réforme du cours en 2020) et de TD;
  - Modélisation et apprentissage statistique;
  - Évaluation par projet et examen final;
  - Collaborateurs: 3 Responsables du cours et 5 autres chargés de TD;
- **Gestion et traitement des données environnementales;**
  - Cursus: Socle commun de Domaine 3;
  - 90 étudiants répartis en 4 groupes de TP;
  - Responsable de cours et chargé de TD;
  - Outils statistiques spécifiques pour les données environnementales (modèles pour les données de présence/absence, comptage et courbes);
  - Évaluation par compte rendu de TP;
  - Collaborateurs: 3 chargés de TD;
- **Outils Mathématiques et statistiques pour la santé;**
  - Cursus: Socle commun de Domaine 4;
  - 96 étudiants répartis en 4 groupes de TP;
  - Responsable de cours et chargé de TD;
  - Outils statistiques spécifiques pour les données de santé (modèles pour les données de présence/absence, analyse de survie, modèles non linéaires, modèles SIR d'épidémiologie);
  - Évaluation par compte rendu de TP et examen écrit;
  - Collaborateurs: 3 chargés de TD;
- **Analyse statistique de données environnementales;**
  - Cursus: UC à choix (UC4);
  - Une dizaine d'étudiants;
  - Intervenant (cours et TP);
  - Modélisation hiérarchique bayésienne;
  - Évaluation par compte rendu de TP;
  - Collaborateurs: 3 autres intervenants;
- **Ingénierie par la simulation numérique;**
  - Cursus: UV projet à choix;
  - Une dizaine d'étudiants;

- Intervenant (cours et TP);
- Modèles d'équations différentielles stochastiques, simulation pour le calcul d'intégrales;
- Évaluation par projet;
- Collaborateurs: 5 autres intervenants;

## **Interventions dans le cursus ingénieur 3A**

- **Analyse statistique de données;**
  - Cursus: Dominante BIOTECH du cursus Agroparistech;
  - Une dizaine d'étudiants;
  - Responsable de cours et de TP;
  - Analyse statistique *end to end* (du nettoyage de données à l'analyse);
  - Évaluation par projet;
  - Collaboratrices: 2 responsables, généticiennes;

### **2.1.1.1 Interventions dans le Master Biologie Ecologie Evolution**

- **Modèles aléatoires pour l'écologie et l'évolution;**
  - Cursus: UC à choix du master BEE;
  - Une dizaine d'étudiants;
  - Responsable de cours et de TP;
  - Modèles probabilistes Markoviens;
  - Évaluation par rendus de TP;
  - Collaborateurs: 2 intervenants;
- **Modèles statistiques en écologie;**
  - Cursus: UC à choix (mais fortement recommandée) du master BEE ;
  - 60 étudiants répartis en 2 groupes de TP;
  - Chargé TP;
  - Modèles statistiques et techniques d'inférence en écologie;
  - Évaluation par projet et examen écrit;
  - Collaborateurs: 2 intervenants en cours et un autre chargé de TP;
- **Introduction à la modélisation bayésienne en écologie;**
  - Cursus: UC à choix du master BEE;
  - 8 étudiants;
  - Intervenant;



- Introduction à la modélisation bayésienne (modèles et inférence);
- Évaluation par compte rendu de TP;
- Collaborateurs: 2 intervenants;

#### **2.1.1.2 Interventions dans le Master d'Ecologie Tropicale de Kourou**

- **Outils Mathématiques pour l'analyse de données;**
  - Cursus: UC obligatoire du master d'écologie Tropicale;
  - 12 étudiants;
  - Responsable de cours;
  - Introduction aux statistiques opérationnelles en écologie;
  - Évaluation par examen écrit;

### **2.1.2 Démarches pédagogiques, responsabilités assumées**

#### **2.1.2.1 Responsabilités**

- Responsable des statistiques pour:
  - Socle commun de domaine 3 des 2A Agroparistech (à peu près 90 étudiants);
  - Socle commun de domaine 4 des 2A Agroparistech (96 étudiants);
  - Dominante Biotech 3A Agroparistech (10 étudiants);

#### **2.1.2.2 Démarches pédagogiques**

- Suivi de la formation à la pédagogie d'Agroparistech en 2020-2021 (120 heures de formation);

### **2.1.3 Création - Innovation - Evolution - Actualisation – Prospective**

#### **2.1.4 Encadrement d'étudiants pour l'enseignement (hors activités de recherche)**

- Encadrement en 2021-2022 et 2022-2023 de deux étudiants du Master MathSV en projet annuel;

- Suivi de CEI de divers étudiants Agroparistech:
  - Mathilde Vuillemin;
  - Jérémie Sénut;
  - Hélène Philippe;
- Suivi de stage pour divers stages d'étudiants Agroparistech de 2A.

## **2.2 Participation à l'enseignement en France hors de l'établissement**

- **R avancé pour la data-science;**
  - 18h entre 2019 et 2020;
  - Cursus: Formation continue en Data Science (Paris Sciences et lettres);
  - 12 étudiants;
  - Responsable;
  - Cours de R avancé pour la manipulation de données et la visualisation;
- **Introduction aux méthodes de Monte Carlo;**
  - 24h en 2020;
  - Cursus ingénieur de l'ENSTA (2ème année);
  - 30 étudiants;
  - Responsable;
  - Introduction aux méthodes de Monte Carlo et à l'inférence bayésienne;

## **2.3 Missions d'enseignement à l'étranger**

## **2.4 Création de documents pédagogiques et publications d'enseignement**

- Vidéos de supports pour les cours de socle commun de D3 et D4 (mise sur ecampus);
- Création d'une autoformation R à destination des étudiants du tronc commun de 2A:
  - Supports `.html`

- Quiz d’autoformation en ligne sur ecampus.
- Mise à disposition des supports de cours (slides, polycopiés, examens corrigés) pour tous mes cours;

# 3 Activités de recherche et de développement

## 3.1 Orientation générale

### Contexte

Depuis 2012, début de ma thèse à IFREMER, mes travaux de recherche se sont focalisés sur deux domaines majeurs de la modélisation statistique, les modèles à variables latentes, et les processus définis par une équation différentielle stochastique. Ma démarche scientifique consiste à partir de données et d'une question biologique, et formaliser un modèle statistique dont l'expression fait le lien entre les données observées et des inconnues. L'estimation de ces inconnues est alors censée apporter une réponse à la question biologique initiale. Ainsi, la capacité à faire de l'estimation dans le modèle posé est cruciale dans cette démarche statistique. Ceci implique alors de i) proposer un estimateur (une méthode de détermination des inconnues à partir des données), ii) fournir un code d'implémentation, iii) évaluer ses performances pratiques, iv) fournir des garanties théoriques notamment ses propriétés asymptotiques.

Mes travaux de recherche depuis mon recrutement ont perduré dans cette voie, avec un accent mis sur les méthodes d'inférence pour les modèles de Markov caché, notamment dans le cadre de modèles à temps continus imparfaitement observés.

### Inférence dans des modèles de Markov caché généraux: définition d'un nouvel estimateur avec garanties théoriques

- (Gloaguen, Le Corff, and Olsson 2022) *A pseudo-marginal sequential Monte Carlo online smoothing algorithm*;
- (Martin et al. 2023) *Backward importance sampling for online estimation of state space models*;

Depuis leur introduction dans les années 60, les HMMs ont fait l'objet d'une grande quantité d'études, cette popularité s'expliquant par leur nombre important d'applications dans des domaines aussi variés que la localisation, la poursuite d'objets en environnement complexe, la restauration de signaux, et comme nous le verrons dans la section suivante, la détection de régimes d'activités. Formellement est une suite d'observations  $Y_{0:T} := Y_0, \dots, Y_T$  (typiquement, les vitesses des individus) ordonnées dans le temps dont les distributions sont définies conditionnellement à la valeur d'un processus de Markov  $X_{0:T} = X_0, \dots, X_T$ . Ce processus **non observé** est caractérisé par sa loi initiale  $\chi_0$  et sa loi de transition  $q(\cdot, \cdot)$ . Dans un cadre paramétrique, les lois d'émissions (loi de  $Y_t|X_t$ ) et de transition (loi de  $X_{t+1}|X_t$ ) dépendent d'une inconnue  $\theta \in \Theta$ , où  $\Theta$  est l'espace des paramètres. On a ainsi le modèle suivant

$$\begin{array}{ll} X_0 \sim \chi_0^\theta(dx_0) & \text{Loi initiale du processus caché} \\ X_t | \{X_{t-1} = x_{t-1}\} \sim q_t^\theta(x_{t-1}, dx_t), \ t \geq 1 & \text{Loi de transition du processus caché} \\ Y_t | \{X_t = x_t\} \sim g_t^\theta(x_t, dy_t), \ t \geq 0 & \text{Loi d'émission des observations} \end{array}$$

Ce formalisme est adapté pour faire de la segmentation de séries temporelles. En halieutique, par exemple, les navires de pêche sont suivis par GPS, fournissant une séquence  $Y_{0:n}$  de vitesses. À partir de ces vitesses, les halieutes souhaitent quantifier l'effort de pêche du navire, *i.e.*, le temps passé à pêcher au cours de son déplacement. Une vision probabiliste du problème consiste à dire que les observations sont des variables aléatoires dont la loi dépend de l'activité du navire (pêche ou non pêche, par exemple), qui est non observée. Cette activité cachée a une dynamique temporelle (l'activité de pêche dure dans le temps), qu'on modélisera par une chaîne de Markov avec un nombre fini d'états. Dans un tel modèle, où l'espace d'état est discret et les lois d'émissions sont paramétrés, l'inférence par maximum de vraisemblance se fait de manière bien connue en utilisant un algorithme itératif, l'algorithme *Expectation Maximization* (EM).

Un cas plus compliqué est celui où le processus caché est à valeurs dans  $\mathbb{R}^{d_x}$ , avec une dynamique potentiellement non gaussienne et non linéaire. Par exemple si on suppose que ce processus est la réalisation d'un processus de diffusion, c'est à dire que:

$$dX_t = f(X_t, t; \theta)dt + \sigma(X_t, t; \theta)dW_t,$$

où  $\{W_t\}_{t \geq 0}$  est un mouvement Brownien standard,  $f(\cdot, \cdot)$  est la *fonction de dérive* et  $\sigma(\cdot, \cdot)$  est la *fonction de diffusion*. La solution d'une EDS, quand elle existe, est un processus stochastique Markovien en temps continu. Cette description en temps continu est attractive en écologie, notamment car elle permet de s'affranchir de problèmes d'échantillonnage. Cependant, faire de l'inférence dans un tel modèle est nettement plus compliqué que dans le cadre discret, car l'étape E de l'algorithme EM n'est plus explicite.

Au cours d’une trilogie d’articles, qui avait commencé en 2018 avec (Gloaguen, Étienne, and Le Corff 2018) et qui vient de s’achever, nous avons pu proposer, avec mes coauteurs et coautrices:

1. Un nouvel algorithme pour faire de l’inférence statistique dans un tel modèle;
2. Des garanties théoriques sur ce nouvel algorithme;
3. Une illustration pratique des performances de cet algorithme.

Cette séquence de travaux s’est faite en collaboration avec Sylvain Le Corff (Télécom Sud-Paris puis Sorbonne Université), Jimmy Olsson (Institut KTH de Stockholm), Marie-Pierre Etienne (Agrocampus-Ouest) et une doctorante, Alice Martin (Ecole polytechnique).

## Problématiques appliquées

- (Michelot et al. 2019) *The Langevin diffusion as a continuous-time model of animal movement and habitat selection*;
- (Gloaguen et al. 2021) *Scalable clustering of segmented trajectories within a continuous time framework: application to maritime traffic data*;
- (Chagneux et al. 2023) *Macrolitter video counting on riverbanks using state space models and moving cameras*;

Les modèles à temps continus et les HMMs sont attractifs dans les applications où l’échantillonnage est irrégulier et bruité, et où les données présentent une structure de dépendance complexe.

Ainsi, les travaux théoriques discutés plus hauts se sont accompagnés d’applications concrètes. Par exemple, avec Théo Michelot, un doctorant en écologie de l’université de Sheffield, nous avons joint la théorie des algorithmes de Monte Carlo par chaîne de Markov (MCMC) avec des approches d’équations différentielles stochastiques guidées par un potentiel pour définir un modèle novateur décrivant le déplacement d’un animal dans un environnement hétérogène. Ce modèle est fourni avec une méthode d’estimation efficace, aux propriétés théoriques connues, et avec une performance pratique testée. Ceci a donné lieu à une publication dans une revue de méthodes pour l’écologie et l’évolution.

D’un point de vue opérationnel, dans un contexte de données massives, le cadre des HMMs pour les trajectoires a aussi été utilisé pour faire du clustering de gros volumes de données, et ainsi de la détection d’anomalies, dans le cadre de la gestion du trafic maritime. Dans ce travail, l’objectif était de déterminer une typologie des trajectoires observées dans des zones de trafic maritime intense, comme le rail d’Ouessant. Ce travail

a donné lieu à une publication dans un numéro spécial d'une revue d'apprentissage statistique, numéro consacré aux traitements de données satellitaires.

Depuis deux ans, je coencadre une thèse dans le cadre du projet [Plastic Origins](#), organisé par l'association Surfrider, dont le but est d'établir une cartographie des déchets plastiques dans les rivières françaises. Le premier objectif de la thèse est de définir, à partir vidéos filmées par des kayakistes le long des rivières, un algorithme de comptage du nombre de déchets dans des vidéos. Il s'agit d'une tâche semi-supervisée car les vidéos elles-mêmes ne sont pas annotées. Nous disposons simplement d'un ensemble d'images indépendantes de ces vidéos, où la présence d'objets est annotée. Le cadre choisi pour cet algorithme consiste à

1. Utiliser des modèles classiques de deep learning pour entraîner un détecteur efficace d'objets plastiques. Ce détecteur sera appliqué séquentiellement sur toutes les images de la vidéo.
2. Considérer les sorties de ces détecteurs comme les observations imparfaites d'un état caché à dynamique Markovienne: l'ensemble des positions des vrais objets.

Cette approche, mêlant deep learning et HMMs, a donné lieu à un premier algorithme de comptage efficace qui sera bientôt mis en production par l'association Surfrider. De plus, cela a donné lieu à une publication récente dans la nouvelle revue de la Société Française de Statistiques: *Computo*, revue de statistique gratuite pour les auteurs, et les lecteurs, consacrée à une science véritablement reproductible.

## Animation de recherche

Depuis mon recrutement, je me suis investi dans différentes actions d'animation de recherche. À l'échelle locale, je suis actuellement coresponsable avec Joon Kwon, membre de mon UMR, de l'organisation du séminaire de Statistiques de mon unité.

Je coanime également, avec Julien Stoeckh (Université Paris Sciences et lettres) et Sylvain Le Corff (Sorbonne Université), un réseau académique parisien sur la statistique bayésienne ([All about that Bayes](#)), qui se réunit mensuellement à Agroparistech. Depuis plus récemment, je coanime avec Guillaume Kon Kam King (INRAe Jouy en Josas) et Nicolas Chopin (ENSAE) un groupe de lecture autour des méthodes de Monte Carlo sur le plateau. Ce groupe de lecture rassemble différents chercheurs du plateau. Cet animation a vocation à créer des collaborations dans notre nouvel environnement de travail qu'est le plateau de Saclay.

À l'échelle nationale, je coanime avec Julien Papaix et Emily Walker (tous deux INRAe Avignon) le réseau INRAe [CisStats](#) (statistics for citizen sciences), qui réunit des chercheurs et des associatifs autour des sciences participatives.

Je me suis également investi dans l'organisation de conférences. J'ai organisé avec Sophie Donnet et Eric Parent, d'Agroparistech, la conférence "Statistiques au sommet de Rochebrune" en Mars 2022, et j'ai organisé avec Julien Stoehr et Sylvain Le Corff la conférence internationale [Bayes at CIRM](#) en Octobre 2021, sur le thème de l'inférence bayésienne *end to end*, i.e. de la théorie aux applications récentes en deep learning.

Depuis 2014, je participe activement au [groupe de recherche Ecologie Statistique](#), regroupant chercheurs en statistiques et en écologie, à l'interface des deux disciplines. Je suis également investi dans le groupe [State of the R](#), animé par des chercheurs de l'équipe SOLSTIS, regroupant des chercheurs et ingénieurs autour des dernières avancées de **R** et **Python** pour les statistiques. De plus, je suis membre de la Société Française de Statistiques, dans laquelle je fais partie du tout nouveau groupe de statistiques bayésiennes.

Enfin, avec Julien Chiquet et Nicolas Jouvin (Agroparistech) et Alline de Reis (Ecole vétérinaire de Maison-Alfort), nous avons en place un *data challenge* sur la plateforme RAMP développée par l'INRIA. L'objectif du *data-challenge* est de faire de la classification de vidéos d'embryons bovins. Cette démarche en partenariat avec une généticienne est l'occasion de lier recherche et enseignement, car le défi est avant tout adresser aux étudiants.

## 3.2 Cadre structurel de la recherche et du développement

Ma recherche s'effectue dans l'unité mixte de recherche (UMR) Agroparistech/INRAe "Mathématiques et informatique appliquées" (MIA-Paris-Saclay). Cette UMR recoupe, peu ou prou, le département de recherche Agroparistech Mathématiques, Modélisation, informatique et physique (MMIP). Au sein de cette UMR, je fais partie de l'équipe *Statistical mOdelling and Learning for environnemenT and Ufe Sciences* (SOLSTIS).

## 3.3 Objectifs - Thème(s) (genèse, état actuel, perspective) - Insertion dans les programmes locaux, régionaux, nationaux, européens et internationaux

Ma recherche est à l'interface entre méthodologie, théories et applications. Elle s'est insérée à merveille dans le cadre de l'équipe SOLSTIS et d'Agroparistech. Un bon exemple de travail (en cours) illustrant cette démarche est une collaboration avec Giacomo Sellaan de l'UMR ECOFOG (Agroparistech Kourou). Dans le cadre de mes enseignements,



j'ai été amené à intervenir dans le Master d'écologie Tropicale d'Agroparistech Kourou. Lors d'un séjour sur place, j'ai rencontré Giacomo Sellan, un botaniste, qui m'a évoqué ses problématiques autour de données recensant la composition spécifique de la forêt de Bornéo.

Les questionnements donnent lieu en ce moment même à une collaboration avec Eric Parent et Achille Thin, deux enseignants chercheurs de SOLSTIS. Cette collaboration s'est traduite par l'encadrement d'un stage d'étudiant de deuxième année Agroparistech (Léo Houairi) et par l'encadrement de projets du master Université Paris-Saclay Mathématiques pour les sciences du vivant.

Au travers du réseau de recherche CISStats, et de la collaboration avec l'association Surfrider, ma recherche s'inscrit également dans une dynamique de sciences citoyennes, afin de réunir ensemble acteurs du monde associatifs et chercheurs en apprentissage statistique.

D'un autre point de vue, comme beaucoup de collègues de ma discipline, mon tropisme a été orienté vers l'apprentissage profond dans les modèles statistiques. Par exemple, la thèse de Mathis Chagneux allie les modèles statistiques connus que sont les HMMs aux modèles d'apprentissage modernes, tels que les réseaux de neurones. Cette alliance des deux est intéressante et laisse augurer une belle progression des résultats théoriques expliquant les performances pratiques des modèles profonds.

L'exemple de l'apprentissage en ligne dans les vidéos est pour cela très parlant, où l'on a besoin d'un côté d'une modélisation adaptée pour des données structurées (ici, des séquences d'images), mais aussi de modèles parcimonieux qui évitent un trop gros coût de calcul, et donc permettent un déploiement opérationnel.

Dans la suite, je développe mes réflexions sur mes perspectives de recherche.

### **3.4 Encadrement de la recherche dans le cadre de l'activité de recherche personnelle**

- *Novembre 2022* - ... Coencadrement du post doctorat de Saint-Clair Chabbert-Liddel sur l'identification de réseaux d'interactions de pêcheurs à partir de leurs suivis GPS;
- *Novembre 2020* - ... Coencadrement de la thèse de Mathis Chagneux sur l'apprentissage en ligne à partir de données structurées, avec un focus applicatif de comptage de déchets plastiques dans des vidéos;
- *Mars - Juin 2022* Encadrement du stage de Odin Ruumianowski (2A Agroparistech) sur l'identification des facteurs de sénescence chez le cerf.

- *Mars - Juillet 2021* Encadrement du stage de Léo Houairi (2A Agroparistech) sur la modélisation jointe de distribution d'espèces dans la forêt de Bornéo;
- *Février - Aout 2020* Encadrement du stage de M2 de Gaspard Baudrin sur l'application du modèle de Langevin en temps continu pour identifier les facteurs environnementaux guidant le déplacement des navires de pêche.

### **3.5 Développement rural, agro-industriel**

### **3.6 Recherche dans le cadre de la coopération internationale**

- Participation au processus d'évaluation pour une demande de financement ANR;
- Participation au processus de relecture d'articles pour des revues à comité de lecture;

### **3.7 Publications scientifiques et valorisation**

Voir Annexe G;

## 4 Autres activités

### 4.1 Activités de prestation de service et d'expertise

### 4.2 Responsabilité, fonctions et réseaux au service de l'établissement (y compris à l'international)

- *Mars 2019* - ... Membre de la commission de titularisation;
- *Mars 2021* - ... Membre de la commission des salles informatiques;

### 4.3 Participation à des jurys (membre ou rapporteur) internes et externes

- *Juin 2023* Membre interne pour un jury de recrutement de Maître(sse) de conférences en statistiques au sein du département MMIP d'Agroparistech;

### 4.4 Responsabilités et fonctions externes à l'établissement

- *2019* -... Coanimation du groupe de recherche francilien *All about that Bayes* de la statistique bayésienne (principaux membres: Sorbonne Université, Paris Sciences et Lettres);
- *2022* -... Coanimation du groupe de recherche *Monte Plateau* regroupant des chercheurs du plateau de Saclay autour des méthodes de Monte Carlo.

## 4.5 Rayonnement national et international

### 4.5.1 Présentations en conférences et ateliers de recherche

- *Octobre 2018* Groupe de recherche écologie statistique (Grenoble);
- *Mai 2019* Assemblée générale du département INRA MIA (Jouy-en-Josas);
- *Juin 2019* Journées de la statistique française (Nancy);
- *Mars 2022* Statistiques au sommet de Rochebrune (Megève);

### 4.5.2 Présentations en séminaire de recherche

- *Janvier 2019* Séminaire de l'Institut Henri Poincaré (Paris);
- *Février 2019* Séminaire de statistiques d'Agroparistech (Paris);
- *Juillet 2019* Séminaire à l'IRISA (Vannes);
- *Mai 2020* Séminaire de statistiques d'Agroparistech (Paris);
- *Juin 2021* Séminaire de statistiques de l'IMAG (Montpellier);
- *Octobre 2021* Séminaire d'ECOFOG (Kourou);
- *Septembre 2022* Séminaire d'hommage à Eric Parent (Paris);
- *Janvier 2023* Séminaire de l'IRMAR (Rennes);
- *Février 2023* Séminaire du laboratoire Raphaël Salem (Rouen);
- *Mars 2023* Séminaire du LMBA (Vannes);

## 4.6 Appartenance à des commissions et des groupes de travail institutionnels

- Membre de la société française de statistique (SFDS);
- Membre fondateur du bureau du groupe “Statistique bayésienne” de la SFDS;

## 5 Réflexion sur les activités et perspectives

- Chagneux, Mathis, Sylvain LeCorff, Pierre Gloaguen, Charles Ollion, Océane Lepâtre, and Antoine Bruge. 2023. “Macrolitter Video Counting on Riverbanks Using State Space Models and Moving Cameras.” *Computo*. <https://doi.org/10.57750/845mf805>.
- Gloaguen, Pierre, Laetitia Chapel, Chloé Friguet, and Romain Tavenard. 2021. “Scalable Clustering of Segmented Trajectories Within a Continuous Time Framework: Application to Maritime Traffic Data.” *Machine Learning*, 1–27.
- Gloaguen, Pierre, Marie-Pierre Étienne, and Sylvain Le Corff. 2018. “Online Sequential Monte Carlo Smoother for Partially Observed Diffusion Processes.” *EURASIP Journal on Advances in Signal Processing* 2018: 1–14.
- Gloaguen, Pierre, Sylvain Le Corff, and Jimmy Olsson. 2022. “A pseudo-marginal sequential Monte Carlo online smoothing algorithm.” *Bernoulli* 28 (4): 2606–33. <https://doi.org/10.3150/21-BEJ1431>.
- Martin, Alice, Marie-Pierre Étienne, Pierre Gloaguen, Le Corff Sylvain, and Jimmy Olsson. 2023. “Backward Importance Sampling for Online Estimation of State Space Models.” *Journal of Computational and Graphical Statistics* 0 (ja): 1–27. <https://doi.org/10.1080/10618600.2023.2174125>.
- Michelot, Théo, Pierre Gloaguen, Paul G. Blackwell, and Marie-Pierre Étienne. 2019. “The Langevin Diffusion as a Continuous-Time Model of Animal Movement and Habitat Selection.” *Methods in Ecology and Evolution* 10 (11): 1894–1907. <https://doi.org/https://doi.org/10.1111/2041-210X.13275>.