

---

## CURRICULUM VITÆ

---

### Gérard BIAU

Né le 26 février 1973.

Nationalité : française.

#### Situation professionnelle

Professeur classe exceptionnelle 2 à Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation (LPSM).

#### Coordonnées personnelles

74 rue Dunois, appartement 32-393, 75013 Paris.

Téléphone : 09 84 44 37 86 ou 06 10 02 71 40.

#### Coordonnées professionnelles

Sorbonne Université

Laboratoire de Probabilités, Statistique et Modélisation

Boîte 158, tour 15-25, bureau 215

4 place Jussieu

75005 Paris, France.

Téléphone : 01 44 27 85 63

Fax : 01 44 27 33 42

e-mail : [gerard.biau@sorbonne-universite.fr](mailto:gerard.biau@sorbonne-universite.fr)

Web : <https://perso.lpsm.paris/~biau>.

### Parcours universitaire

---

**1994-95** Maîtrise de mathématiques pures, Université Montpellier 2.

**1995-97** Elève de l'Ecole des Mines de Paris, option géostatistique et probabilités appliquées. Diplômé Ingénieur civil des Mines.

**1997-98** DEA de biostatistique, option statistique mathématique, Université Montpellier 2.

**2000** Agrégation de mathématiques, option probabilités et statistique.

**1998-00** Doctorat de mathématiques appliquées.

Thèse soutenue à l'Université Montpellier 2 le 15 décembre 2000.

Titre : “*Méthodes itératives en estimation fonctionnelle et systèmes dynamiques*”.

Directeur de thèse : Alain Berlinet, Université Montpellier 2.

Jury : Jan Beirlant, Alain Berlinet, Denis Bosq (rapporteur), Luc Devroye (rapporteur), Gérard Grégoire, Pierre Jacob, Hans Wackernagel.

**2001-04** Maître de conférences à l'Université Paris VI.

- 2003** Habilitation à diriger des recherches.  
Thèse soutenue à l'Université Paris VI le 8 décembre 2003.  
Titre : “*Contribution à la statistique non paramétrique et ses applications*”.  
Jury : Anestis Antoniadis (rapporteur), Alain Berlinet, Lucien Birgé, Denis Bosq, Paul Deheuvels, Luc Devroye, Marc Hallin (rapporteur), Gábor Lugosi (rapporteur).
- 2004-07** Professeur à l'Université Montpellier 2.
- 2007** Mutation en qualité de professeur à Sorbonne Université à compter du 1<sup>er</sup> septembre 2007.
- 2012-17** Membre junior de l'Institut Universitaire de France.

## Enseignements

---

Ma qualité de maître de conférences à l'Université Paris VI (2001-2004), de professeur à l'Université Montpellier 2 (2004-2007) puis à Sorbonne Université (anciennement Université Paris VI, depuis 2007) m'ont conduit à dispenser des enseignements de nature théorique (probabilités, statistique mathématique) ou appliquée (TP informatiques de statistique, encadrement de projets). Ces enseignements, qu'il s'agisse de cours ou de travaux dirigés, ont couvert la quasi-totalité des niveaux universitaires, du L1 au M2 recherche en passant par la préparation à l'agrégation de mathématiques (option probabilités et statistique). Ils ont concerné aussi bien des étudiants de filières spécialisées en mathématiques appliquées que des étudiants non spécialistes (en biologie, notamment). Par ailleurs, j'ai eu l'occasion, au cours d'enseignements complémentaires, de m'adresser à un public d'élèves ingénieurs ou d'école de commerce (Polytech'Montpellier, Ecole des Mines de Paris, EDHEC de Lille). De 2009 à 2014, j'ai effectué la moitié de mon service statutaire à l'Ecole Normale Supérieure de la rue d'Ulm.

Diverses responsabilités pédagogiques m'ont été confiées à l'Université Montpellier 2 et à l'Université Paris VI. En particulier, j'ai été responsable du M1 commun aux deux spécialités mathématiques et statistique du master Informatique-Mathématiques-Statistique de l'Université Montpellier 2 (2004-2007), directeur des études de l'Institut de Statistique de l'Université de Paris (ISUP, 2007-2012) et directeur adjoint du master Mathématiques et Applications de l'Université Paris VI (2011-2013). J'ai assuré la fonction de directeur adjoint du Laboratoire de Probabilités, Statistique et Modélisation (LPSM, 2018-2021) après avoir été directeur du Laboratoire de Statistique Théorique et Appliquée (LSTA, 2013-2018). Je suis depuis 2019 directeur nommé du Sorbonne Center for Artificial Intelligence (SCAI).

## Thèmes de recherche

---

Mes activités de recherche se sont organisées autour de six thèmes privilégiés :

- Systèmes dynamiques et chaos.
- Statistique non paramétrique.
- Statistiques spatiales.
- Méthodes combinatoires et sélection de modèle.
- Apprentissage statistique, courbes et grande dimension.

- Bioinformatique.

J'ai développé les trois premiers thèmes durant ma thèse et mes premières années de maître de conférences. Je me suis ensuite concentré sur le quatrième axe de recherche, qui a été au cœur de mes travaux d'habilitation. Aujourd'hui, je me consacre essentiellement aux deux derniers thèmes, en essayant de donner à ma recherche une coloration située au carrefour de la statistique et de l'algorithmique.

### **Systemes dynamiques et chaos**

- Utilisation des propriétés statistiques de certains systemes dynamiques deterministes chaotiques en vue de mettre au point une nouvelle technique de simulation. En collaboration avec Alain Berlinet (Professeur, Université Montpellier 2). Rapport de recherche [b].
- Etude de la convergence et du comportement dynamique d'estimateurs de la densité au sens des critères de la théorie de l'information. En collaboration avec Jan Beirlant (Professeur, Université Catholique de Leuven), Alain Berlinet et Igor Vajda (Professeur, Institut de Théorie de l'Information et d'Automatique de Prague). Publications [3], [6] et [15].
- Utilisation des méthodes statistiques non paramétriques pour l'estimation des systemes dynamiques chaotiques. Obtention de résultats minimax non asymptotiques via une triangulation de l'espace en cellules de Voronoi. En collaboration avec Alain Berlinet. Publication [4].
- Etablissement d'une classification des transformations chaotiques via le concept de sensibilité aux conditions initiales. Clarification du lien entre la positivité de l'exposant de Lyapunov et le comportement chaotique. En collaboration avec Christophe Abraham (Professeur, Montpellier SupAgro) et Benoît Cadre (Professeur, Université Rennes 2). Publications [5] et [12].

### **Statistique non paramétrique**

- Etude d'algorithmes itératifs associés à l'estimation fonctionnelle sur échantillon. Application à la sélection du paramètre de lissage d'un estimateur à noyau par plug-in itéré. Publication [1].
- Rédaction d'un chapitre intitulé "*Estimation de densité et prise de décision*" paru dans le livre *Décision et Reconnaissance de Formes en Signal*. En collaboration avec Alain Berlinet. Publication [c].
- Analyse des propriétés asymptotiques d'un estimateur simple et facilement calculable du mode d'une densité multivariée. En collaboration avec Christophe Abraham et Benoît Cadre. Publications [8] et [11].
- Définition de l'indice de régularité de la dérivée symétrique d'une mesure de probabilité. Mise au point et étude asymptotique d'un estimateur non paramétrique de cet indice inspiré de l'estimateur de Pickands. Application à la sélection du nombre optimal de voisins pour l'estimateur des  $k$ -plus proches voisins de la densité. En collaboration avec Jan Beirlant et Alain Berlinet. Publication [30].
- Etude d'un nouveau test non paramétrique relatif à l'homogénéité de deux échantillons. Analyse des propriétés de grandes déviations et de la normalité asymptotique de la statistique de test. En collaboration avec László Györfi (Professeur, University of Technology and Economics de Budapest). Publications [22] et [d].

- Obtention, à partir d’outils de la géométrie riemannienne, des vitesses de convergence exactes (pour la différence symétrique) de l’estimateur à boules du support d’une densité de probabilité multivariée. En collaboration avec Benoît Cadre et Bruno Pelletier (Professeur, Université Rennes 2). Publication [32].
- Définition et étude asymptotique d’un nouvel estimateur (construit sur un graphe) du nombre de clusters d’une distribution de probabilité. En collaboration avec Benoît Cadre et Bruno Pelletier. Publication [26].
- Mise au point et analyse des propriétés asymptotiques de deux nouvelles stratégies non paramétriques de prévision de séries temporelles, que l’on suppose seulement stationnaires et ergodiques. Ces méthodes reposent sur l’agrégation convexe de prédicteurs non paramétriques élémentaires, appelés “experts” dans ce contexte. En collaboration avec Kevin Bleakley (Chargé de recherche, Inria Saclay – Ile-de-France), László Györfi et György Ottucsák (Chercheur, University of Technology and Economics de Budapest). Publication [38].
- Extension des techniques d’agrégation précédentes au cas de la prévision quantile. En collaboration avec Benoît Patra (chief technology officer, société iAdvize). Publication [42].
- Etablissement de la normalité asymptotique de l’estimateur à boules du support d’une densité de probabilité multivariée. Cet article répond, au passage, à une question laissée ouverte par Peter Hall en 1985 dans un article sur l’analyse de la *vacancy*. En collaboration avec Benoît Cadre, David Mason (Professeur, Université du Delaware à Newark) et Bruno Pelletier. Publication [36].
- Etude asymptotique complète (convergence, vitesses de convergence, approximation forte, théorèmes limites, optimalité) d’un nouvel estimateur de la densité de type  $k$ -plus proches voisins à poids. L’analyse de cet estimateur, qui généralise l’estimateur des  $k$ -plus proches voisins traditionnel, est motivée (et illustrée sur des données synthétiques et réelles) par des problématiques d’inférence géométrique et topologique. En collaboration avec Frédéric Chazal (Directeur de recherche, Inria Saclay – Ile-de-France), David Cohen-Steiner (Chargé de recherche, Inria Sophia Antipolis – Méditerranée), Luc Devroye (Professeur, Université McGill de Montréal) et Carlos Rodríguez (Professeur, State University of New York à Albany). Publication [43].
- Analyse asymptotique fine, au sens du critère de proximité de Pitman, de différents estimateurs dits “shrinkés” de la variance et de l’écart-type. En collaboration avec Yannis Yatracos (Professeur, Tsinghua University de Pékin). Publication [47].
- Mise au point et preuve de la convergence d’un estimateur de la fonction de régression multivariée invariant par transformation affine des données. Cet estimateur est basé sur une technique de plus proches voisins évalués à l’aide d’une métrique empirique. En collaboration avec Luc Devroye, Vida Dujmović (Professeur, Université Carleton d’Ottawa) et Adam Krzyżak (Professeur, Université Concordia de Montréal). Publication [48].
- Analyse mathématique approfondie de l’algorithme de simulation bayésien ABC (pour *Approximate Bayesian Computation*). Ce travail met, d’une part, en évidence le lien étroit entre l’approche ABC et les méthodes de type plus proches voisins et propose, d’autre part, un nouvel estimateur de la densité conditionnelle, hybride entre un estimateur des  $k$ -plus proches voisins et un estimateur à noyau. En collaboration avec Frédéric Céro (Chargé de recherche, Inria Rennes – Bretagne Atlantique) et Arnaud Guyader (Professeur, Sorbonne Université). Publication [52].

- Obtention de la loi asymptotique d’une statistique récemment utilisée dans des problèmes de détection de ruptures. En nous appuyant sur cette loi asymptotique, nous proposons un nouveau protocole de test, mis en œuvre au travers d’un algorithme rapide et performant. Cet algorithme peut être utilisé sur des signaux de taille importante, comme en attestent plusieurs exemples d’application sur données réelles et simulées. En collaboration avec Kevin Bleakley et David Mason. Publication [58].

## Statistiques spatiales

- Modélisation géostatistique du problème dit du *downscaling*. Application à la prévision de précipitations par krigeage de la hauteur d’eau en des points d’estimation “synthétiques” simulés par un modèle de circulation atmosphérique. En collaboration avec Hans von Storch (Professeur, Helmholtz-Zentrum Geesthacht, Allemagne), Hans Wackernagel (Chercheur, Ecole des Mines de Paris) et Eduardo Zorita (Chercheur, Helmholtz-Zentrum Geesthacht). Publications [2] et [a].
- Estimation non paramétrique de la densité dans les champs aléatoires. Etude des différents types de vitesses asymptotiques pouvant être obtenus dans le cas des champs portés par un indice discret, continu et échantillonné. Publications [7] et [10].
- Mise au point et étude asymptotique d’une méthode d’interpolation non paramétrique pour les champs aléatoires. En collaboration avec Benoît Cadre. Publication [16].

## Méthodes combinatoires et sélection de modèle

- Etude minimax non asymptotique de la classe des applications bloc-décroissantes multivariées. Présentation d’un estimateur adaptatif minimax-optimal obtenu à l’aide des techniques de sélection combinatoires développées par Luc Devroye et Gábor Lugosi (Professeur, Université Pompeu Fabra de Barcelone). En collaboration avec Luc Devroye. Publication [9].
- Développement de nouvelles techniques de sélection pour des modèles de type mélanges de densités. Application à des problèmes de tests. En collaboration avec Benoît Cadre, Luc Devroye et László Györfi. Publications [13], [14], [17] et [31].
- Sélection combinatoire des paramètres de lissage pour des estimateurs de la densité de type histogrammes modifiés. En collaboration avec Alain Berlinet et Laurent Rouvière (Maître de conférences, Université Rennes 2). Publication [18].
- Problème de la sélection par méthode combinatoire d’une fonction copule dans une famille paramétrique donnée satisfaisant des propriétés de type Vapnik-Chervonenkis. En collaboration avec Marten Wegkamp (Professeur, Université Cornell). Publication [19].
- Etude des propriétés à distance finie d’estimateurs à noyau automatiques possédant un paramètre de lissage fonctionnel (variant en fonction du point d’estimation et/ou des points de l’échantillon). En collaboration avec Alain Berlinet et Laurent Rouvière. Publication [21].

## Apprentissage statistique, courbes et grande dimension

- Etude des propriétés asymptotiques de la règle du noyau pour la classification supervisée dans des espaces fonctionnels. En collaboration avec Christophe Abraham et Benoît Cadre. Publication [23].

- Développement et étude mathématique d’une méthode de classification automatique de courbes basée sur des développements en séries de Fourier. Application sur données réelles à un problème de reconnaissance vocale. En collaboration avec Florentina Bunea (Professeur, Université Cornell) et Marten Wegkamp. Publication [20].
- Extension de la méthode précédente à des bases d’ondelettes. En collaboration avec Alain Berlinet et Laurent Rouvière. Publication [33].
- Modélisation et analyse asymptotique du problème de la reconstruction statistique d’un graphe évolutif dont les sommets et les arêtes sont aléatoires. En collaboration avec Kevin Bleakley. Publication [24].
- Obtention de nouvelles bornes non-asymptotiques pour le risque de la procédure de clustering dite des  $k$ -means dans un espace de Hilbert séparable. En collaboration avec Luc Devroye et Gábor Lugosi. Publication [29].
- Développement d’une nouvelle méthodologie statistique permettant de prévoir la production manufacturière à partir des réponses individuelles des chefs d’entreprises à l’enquête de conjoncture Industrie de l’INSEE. L’approche est fondée sur des techniques non paramétriques de type forêts aléatoires de Leo Breiman. En collaboration avec Olivier Biau (Administrateur de l’INSEE) et Laurent Rouvière. Publication [28].
- Définition précise et étude des propriétés de convergence des règles de classification dites “randomisées”, construites en effectuant un vote majoritaire sur une collection de règles élémentaires dépendant d’un paramètre aléatoire. Analyse, dans ce contexte, du comportement asymptotique de différentes versions de l’algorithme des forêts aléatoires de Leo Breiman. En collaboration avec Gábor Lugosi et Luc Devroye. Publication [34].
- Analyse des propriétés asymptotiques des *Layered Nearest Neighbours* (LNN, il s’agit de plus proches voisins construits à partir d’une métrique géométrique adaptative) et de leurs connexions avec les forêts aléatoires et l’algorithme *bagging* (*bootstrap-aggregating*). En collaboration avec Luc Devroye. Publication [41].
- Etude complète (définition, convergence, vitesses de convergence) du modèle statistique des  $k$ -plus proches voisins pour les systèmes dits “de recommandation collaborative”. Ces systèmes informatiques, qui sont utilisés par de nombreux sites internet, proposent des suggestions personnalisées aux internautes (livres, films, musique, etc.) évaluées en fonction de la proximité entre leurs goûts. En collaboration avec Benoît Cadre et Laurent Rouvière. Publication [39].
- Identification des vitesses de convergence de l’estimateur des  $k$ -plus proches voisins lorsque les observations prennent leurs valeurs dans un espace fonctionnel. Caractérisation de ces vitesses dans le cas des espaces de Sobolev, des espaces de Besov et des espaces de Hilbert à noyau autoreproduisant. En collaboration avec Frédéric Cérou et Arnaud Guyader. Publication [40].
- Analyse des vitesses de convergence et des propriétés minimax sur des classes de Hölder de l’estimateur du plus proche voisin “baggé”. En collaboration avec Frédéric Cérou et Arnaud Guyader. Publication [37].
- Approfondissement, dans un contexte de sparsité, des propriétés statistiques d’un modèle de forêts aléatoires proposé par Leo Breiman en 2004. Ce travail met en évidence le fait que la vitesse de convergence de l’algorithme ne dépend que de la dimension intrinsèque

du problème (c'est-à-dire le nombre de variables informatives) et non pas de la dimension ambiante. Publication [46].

- Obtention de résultats généraux justifiant l'utilisation conjointe de l'analyse en composantes principales et des estimateurs à noyau (en densité ou régression), dans le cadre de données fonctionnelles ou en grande dimension. En collaboration avec André Mas (Professeur, Université Montpellier 2). Publication [45].
- Analyse, en situation de parcimonie, des propriétés non asymptotiques (inégalités oracles) d'un nouvel estimateur du modèle de régression single-index. Mise en œuvre et test de la méthode sur données réelles et simulées à l'aide d'un algorithme de type *reversible jump Markov chain Monte Carlo*. En collaboration avec Pierre Alquier (Professeur, ESSEC Business School). Publication [49].
- Mise en œuvre d'une procédure de choix de modèle et obtention d'inégalités oracles pour la sélection des paramètres des courbes principales. En collaboration avec Aurélie Fischer (Maître de conférences, Université Paris Cité). Publication [44].
- Définition et étude asymptotique d'une nouvelle famille de règles de classification dites "cellulaires". Dans ce paradigme, les décisions sont fondées sur des arbres dont chaque cellule travaille en toute autonomie, indépendamment de la structure globale de l'arbre, ouvrant ainsi la voie à des architectures parallèles. En collaboration avec Luc Devroye. Publications [50] et [e].
- Invention, étude mathématique et tests intensifs sur données synthétiques et réelles d'une stratégie non linéaire d'agrégation d'estimateurs. Cette méthode originale, dont les performances pratiques sont remarquables, s'accompagne d'un packaging disponible sur le site du CRAN. En collaboration avec Aurélie Fischer, Benjamin Guedj (Chargé de recherche, Inria Lille – Nord Europe) et James Malley (Chercheur, National Institute of Health de Bethesda). Publication [56].
- Etude théorique complète (inégalité oracle, convergence, vitesse) d'une méthode de classification supervisée de trajectoires issues de processus de Cox. La stratégie de classification repose sur une minimisation contrainte d'une version convexe du risque empirique. En collaboration avec Benoît Cadre et Quentin Paris (Assistant professor, National Research University "Higher School of Economics" de Moscou). Publication [55].
- Soit  $\mathbf{X} = (X_1, \dots, X_d)$  un vecteur aléatoire dont les composantes sont indépendantes et identiquement distribuées. Nous analysons dans un premier temps les propriétés asymptotiques surprenantes (convergence, loi limite, etc.) de la  $p$ -norme  $\|\mathbf{X}\|_p = (\sum_{j=1}^d |X_j|^p)^{1/p}$  lorsque la dimension  $d$  tend vers l'infini. Nous étudions dans un second temps le comportement limite de la  $p$ -norme maximale et de la  $p$ -norme minimale associées à un échantillon  $\mathbf{X}_1, \dots, \mathbf{X}_n$  issu de  $\mathbf{X}$ . En collaboration avec David Mason. Publication [53].
- Preuve de la convergence de l'algorithme original des forêts aléatoires dans le cadre des modèles de régression additifs. L'approche repose sur une étude fine des propriétés des coupures d'arbres opérées par la procédure CART (*Classification and Regression Trees*). En collaboration avec Erwan Scornet (Maître de conférences, Ecole Polytechnique) et Jean-Philippe Vert (chief R&D officer, société Owkin, Professeur, Mines Paris - PSL). Publication [54].
- Etude des propriétés asymptotiques d'un algorithme d'estimation de la régression en ligne, distribué et asynchrone. La coopération entre processeurs s'effectue par passage de messages

et, sous des hypothèses convenables, l’architecture conduit à un consensus sur l’estimation. La procédure s’accompagne d’un logiciel en accès libre permettant le passage à l’échelle et mettant en évidence les excellentes performances de l’approche, tant du point de vue des temps de calcul que des résultats statistiques. En collaboration avec Ryad Zenine (senior data scientist, société Contentsquare). Publication [60].

- Revue de la littérature portant sur l’algorithme des forêts aléatoires. L’accent est mis sur les modèles mathématiques existant à ce jour et leur connexion avec les méthodes à noyau, sur le rôle des différents paramètres (nombre de données dans les feuilles des arbres, rééchantillonnage, etc.), et sur les critères connexes de sélection de variables. L’article s’achève par une revue des nombreuses extensions de l’algorithme. En collaboration avec Erwan Scornet. Publication [57].
- Définition et analyse d’un algorithme d’inférence collaborative de la moyenne inconnue  $\theta$  d’une variable aléatoire. Dans ce modèle, un certain nombre d’unités de calcul distribuées en réseau participent à l’estimation séquentielle de  $\theta$  en échangeant des informations via une matrice stochastique  $A$ . Nous donnons des conditions précises sur  $A$  sous lesquelles la performance statistique de chaque unité est comparable à celle d’une entité virtuelle centralisée. Nous montrons en particulier le rôle fondamental joué par les valeurs propres non triviales de  $A$  ainsi que par les graphes expanseurs de Ramanujan, qui fournissent des résultats remarquables pour un coût algorithmique modeste. En collaboration avec Kevin Bleakley et Benoît Cadre. Publication [59].
- Une forêt aléatoire peut être réinterprétée comme une collection de réseaux neuronaux dont les connexions et les paramètres dépendent des arbres de la forêt. Nous exploitons cette analogie pour définir deux estimateurs de la régression hybrides, que nous appelons “forêts aléatoires neuronales”, combinant la simplicité des forêts et la flexibilité des réseaux. Nous établissons la convergence de ces estimateurs et mettons en évidence leurs excellentes performances sur données réelles et simulées. En collaboration avec Erwan Scornet et Johannes Welbl (research scientist, société DeepMind). Publication [62].
- Le *gradient boosting*, très populaire dans le domaine du *machine learning*, est un mécanisme algorithmique permettant de résoudre efficacement des tâches complexes de régression et de classification supervisée. Le principe général consiste à construire progressivement un modèle additif, en agréant séquentiellement des prédicteurs très simples, appelés *weak learners* (typiquement des arbres de décision peu profonds). Malgré des performances pratiques exceptionnelles, la théorie de ces algorithmes est encore fort incomplète et leurs mécanismes mathématiques mal compris. Dans cet article, nous appréhendons le *gradient boosting* sous l’angle de l’optimisation numérique et établissons des propriétés de convergence nouvelles. Nous montrons également que les prédicteurs “boostés” possèdent, sous des conditions adaptées, de bonnes propriétés de convergence statistique, sans qu’il soit nécessaire de recourir à des techniques de régularisation par *early stopping*. En collaboration avec Benoît Cadre. Publication [69].
- Nous présentons une nouvelle version de l’algorithme de *gradient boosting* de Friedman intégrant un schéma de Nesterov accéléré. Nous montrons au travers d’une étude empirique complète que cet algorithme, qui utilise des arbres de décision comme *weak learners*, conserve les excellentes propriétés prédictives du *gradient boosting*. Il est cependant beaucoup moins sensible au choix du paramètre de descente et fournit des prédicteurs considérablement plus parcimonieux en le nombre d’arbres, ce qui constitue un avantage décisif en situation d’apprentissage à grande échelle. En collaboration avec Benoît Cadre et Laurent Rouvière. Publication [61].



- Les *Generative Adversarial Networks* (GANs) sont une classe d’algorithmes génératifs connus pour produire des échantillons factices à la pointe de la technologie. Ces algorithmes utilisent deux réseaux neuronaux (générateur et discriminateur) qui s’améliorent mutuellement via l’optimisation d’un critère “adversarial”. Nous proposons dans cet article une analyse théorique de ce critère, en étudiant certaines de ses propriétés mathématiques et statistiques. Nous approfondissons en particulier le lien étroit unissant les GANs et la divergence de Jensen-Shannon, ainsi que certaines caractéristiques optimales du problème. D’un point de vue plus statistique, nous démontrons des résultats de convergence relatifs au modèle estimé ainsi qu’un théorème central limite. En collaboration avec Benoît Cadre, Maxime Sangnier (Maître de conférences, Sorbonne Université) et Ugo Tanielian (senior AI engineer, société Gleamer). Publication [63].
- Les algorithmes d’apprentissage modernes, comme les forêts aléatoires ou les réseaux de neurones, sont souvent qualifiés de “boîtes noires” en raison du nombre élevé et de la complexité des opérations impliquées dans leurs mécanismes internes. Ce défaut d’interprétabilité représente une limitation importante dans les applications impliquant des décisions critiques, en particulier dans les processus de production de l’industrie manufacturière. Dans un tel contexte, les modèles doivent être interprétables, c’est-à-dire simples, stables et prédictifs. Nous attaquons le problème en développant SIRUS (*Stable and Interpretable Rule Set*), un nouvel algorithme de classification supervisée basé sur des forêts aléatoires qui prend la forme d’une courte liste de règles. Alors que les modèles simples sont généralement instables, ce n’est pas le cas de SIRUS, qui hérite en outre d’une qualité prédictive proche de celle des forêts combinée à la simplicité des arbres de décision. L’ensemble de ces propriétés est évalué d’un point de vue théorique et empirique, au travers d’expériences numériques approfondies issues de notre implémentation logicielle `sirus` en R/C++. En collaboration avec Clément Bénard (Ingénieur de recherche, société Safran), Sébastien Da Veiga (expert senior statistique et optimisation, société Safran) et Erwan Scornet. Publication [64].
- Dans l’esprit de la publication [63], nous contribuons à une meilleure compréhension des propriétés mathématiques des Wasserstein GANs (WGANs), une approche cousine des GANs connue pour être plus stable lors de la phase d’entraînement. Nous montrons tout d’abord comment l’architecture des WGANs s’inscrit naturellement dans le contexte des *Integral Probability Metrics* paramétrées par des réseaux de neurones. Cela nous permet en particulier de souligner quelques propriétés d’optimisation intéressantes découlant de l’utilisation d’un discriminateur paramétrique 1-Lipschitz. Dans une optique plus statistique, nous étudions ensuite la convergence des WGANs et clarifions les effets antagonistes du générateur et du discriminateur via des propriétés de compromis. Enfin, l’ensemble est illustré par des expériences utilisant des données synthétiques et réelles. En collaboration avec Maxime Sangnier et Ugo Tanielian. Publication [68].
- Extension de SIRUS au cas de la régression. En collaboration avec Clément Bénard, Sébastien Da Veiga et Erwan Scornet. Publication [66].
- Motivés par la problématique des *Heterogeneous Treatment Effects*, nous présentons de nouvelles perspectives en inférence causale en proposant des variantes des forêts aléatoires adaptées à l’estimation des lois conditionnelles. Pour ce faire, nous reformulons le critère de coupe original de Breiman en termes de distances de Wasserstein entre mesures empiriques, ce qui fournit au passage une extension naturelle de l’algorithme à des sorties multivariées. Des expériences numériques montrent comment notre approche peut contribuer à mener une inférence causale plus transparente dans des situations complexes. En collaboration avec Qiming Du (research associate, société Capital Fund Management), François Petit (Maître

de conférences, Université Paris Cité) et Raphaël Porcher (Professeur-Praticien hospitalier, Université Paris Cité). Publication [67].

- Les avancées pratiques les plus récentes en apprentissage profond, notamment dans le domaine des attaques adversariales et des WGANs, montrent l'importance d'utiliser des réseaux de neurones dont les constantes de Lipschitz sont contraintes. Partant de ce constat, nous travaillons dans cet article à une meilleure compréhension théorique de l'approximation des fonctions 1-Lipschitz par des réseaux de neurones. Nous montrons en particulier, via des résultats de densité, de profondeur et de taille, que des réseaux admettant des activations de type GroupSort sont bien adaptés au problème. L'efficacité des réseaux GroupSort par rapport aux réseaux ReLU, plus classiques mais moins à même d'estimer correctement des fonctions Lipschitz, est illustrée dans une série d'expériences numériques. En collaboration avec Maxime Sangnier et Ugo Tanielian. Publication [65].
- Les *Shapley effects*, qui permettent de mesurer l'importance des variables dans un problème d'apprentissage, sont aujourd'hui très largement utilisés pour interpréter les ensembles d'arbres et les réseaux neuronaux. Leur estimation est cependant un problème difficile, en raison de la complexité du calcul d'une part et des espérances conditionnelles sous-jacentes d'autre part. Par conséquent, les algorithmes actuels permettant de remplir cette tâche présentent de nombreux défauts, au premier rang desquels un temps d'exécution coûteux ou un biais lorsque les variables sont dépendantes. Dans cet article, nous introduisons SHAFF (pour *SHApley eFfects via random Forests*), un algorithme d'estimation rapide et précis des *Shapley effects*. Nous démontrons l'efficacité de SHAFF via une analyse théorique de sa convergence et mettons en évidence ses excellentes performances par rapport aux concurrents au travers d'expériences numériques approfondies. En collaboration avec Clément Bénard, Sébastien Da Veiga et Erwan Scornet. Publication [71].
- En nous appuyant sur l'interprétation d'un réseau neuronal récurrent (RNN) comme la discrétisation d'une équation différentielle ordinaire en temps continu, nous montrons, sous des hypothèses adaptées, que la solution d'un RNN peut être considérée comme une fonction linéaire de la signature du processus d'entrée. Cette connexion nous permet alors d'envisager un RNN comme une méthode à noyau dans un espace de Hilbert à noyau reproduisant approprié. En suivant, nous obtenons des garanties théoriques en termes de généralisation et de stabilité pour une grande classe de réseaux récurrents. En collaboration avec Adeline Fermanian (chercheuse en machine learning, société Califrais), Pierre Marion (postdoctorant, EPFL) et Jean-Philippe Vert. Publication [70].
- Intéressés par la caractérisation géométrique des distributions générées par les WGANs, nous proposons une analyse approfondie de ces derniers, en régime fini et asymptotique. Nous montrons en particulier que les WGANs optimaux sont étroitement liés à des chemins minimisant la somme des carrés des distances euclidiennes entre les points de l'échantillon. Nous mettons également en évidence le fait que les WGANs approchent asymptotiquement la loi cible (au sens de la distance 1-Wasserstein), avec une vitesse de convergence spécifiée et à condition que la famille de fonctions génératives croisse de manière appropriée. Nous démontrons en passant de nouveaux résultats en théorie du transport optimal dans le cadre semi-discret. En collaboration avec Benoît Cadre, Nicolas Klutchnikoff (Maître de conférences, Université Rennes 2), Arthur Stéphanovitch (doctorant, Université Paris Cité) et Ugo Tanielian. Publication [73].
- Les réseaux neuronaux résiduels (ResNets) profonds sont reconnus pour les résultats exceptionnels qu'ils obtiennent dans des tâches d'apprentissage automatique complexes. Les performances remarquables de ces architectures reposent cependant sur une procédure

d'apprentissage délicate, qui doit être soigneusement mise en œuvre afin d'éviter que les gradients ne s'annulent ou n'explorent, en particulier lorsque la profondeur  $L$  augmente. Il n'existe pas, à ce jour, de consensus clair sur la façon de régler cette difficulté, bien qu'une stratégie classique consiste à multiplier la sortie de chaque couche par un facteur d'échelle  $\alpha_L$ . Nous montrons dans un cadre probabiliste qu'avec des initialisations i.i.d. standards, la seule dynamique intéressante se produit pour  $\alpha_L = 1/\sqrt{L}$ , tous les autres choix conduisant à des comportements moins pertinents pour l'apprentissage. Ce facteur d'échelle correspond dans la limite temps continu à une équation différentielle stochastique, contrairement à une interprétation répandue selon laquelle les ResNets profonds sont des discrétisations d'équations différentielles ordinaires (dites neuronales). Pour ce dernier régime, la stabilité est obtenue avec des initialisations corrélées bien spécifiques et  $\alpha_L = 1/L$ . Au final, notre analyse suggère une interaction forte entre le facteur d'échelle  $\alpha_L$  et la régularité des poids en fonction de l'indice des couches. En collaboration avec Adeline Fermanian, Pierre Marion et Jean-Philippe Vert. Publication [75].

- La signature d'un processus représente ce dernier sous la forme d'une suite infinie d'intégrales itérées. Nous proposons dans cette contribution une méthode efficace pour inverser la signature, c'est-à-dire pour reconstruire le processus à partir de l'information de sa signature (tronquée). Après avoir décrit notre méthode, nous établissons sa convergence pour des chemins linéaires par morceaux, accompagnée d'un algorithme. Cet algorithme est parallélisé, ce qui signifie qu'il est adapté pour inverser simultanément un ensemble de signatures. En collaboration avec Adeline Fermanian, Jiawei Chang (Université d'Oxford) et Terry Lyons (Professeur, Université d'Oxford). Publication [72].
- Les réseaux neuronaux informés par la physique (PINNs en anglais, pour *physics informed neural networks*) combinent la qualité prédictive des réseaux et l'interprétabilité des modèles physiques. Les PINNs ont démontré leur utilité pratique dans la résolution d'équations aux dérivées partielles et dans des scénarios de modélisation hybride, où une information de nature physique complète l'échantillonnage aléatoire. Nous montrons dans ce travail que l'apprentissage classique des PINNs souffre d'un *overfitting* systématique. Ce défaut peut être corrigé en ajoutant une régularisation *ridge* au risque empirique, qui garantit la convergence du risque pour des systèmes différentiels linéaires et non linéaires. Toutefois, la convergence forte (au sens des fonctions) des PINNs nécessite une analyse plus approfondie, mobilisant des outils de l'analyse fonctionnelle et du calcul des variations. Pour les systèmes linéaires, une régularisation de type Sobolev (facilement implémentable) conduit à une solution combinant à la fois garantie statistique et cohérence avec la physique sous-jacente. En collaboration avec Claire Boyer (Maître de conférences, Sorbonne Université) et Nathan Doumèche (doctorant, Sorbonne Université et EDF). Publication [76].
- Malgré une utilisation répandue, le lien entre les réseaux neuronaux résiduels et les ODEs neuronales (leur analogue continu) manque encore de fondements mathématiques solides. Nous montrons dans cet article qu'il existe une régularisation implicite des réseaux résiduels profonds vers les ODEs neuronales pour des réseaux non linéaires entraînés par flot de gradient. Nous établissons en particulier que si le réseau est initialisé comme la discrétisation d'une ODE neuronale, alors cette discrétisation reste valable tout au long de la séquence d'apprentissage. Nos résultats s'appliquent pour un temps d'apprentissage fini ou infini, à condition dans ce dernier cas que le réseau satisfasse une condition de type Polyak-Łojasiewicz. De façon notable, cette condition est réalisée pour une famille de réseaux résiduels dont les résidus sont des perceptrons à deux couches, avec un surparamétrage en largeur qui n'évolue que linéairement. Elle implique alors la convergence du flot de gradient vers un minimum global de la perte. En collaboration avec Pierre Marion, Michael Sander

(doctorant, Ecole Normale Supérieure de la rue d'Ulm) et Yu-Han Wu (stagiaire, Ecole Normale Supérieure de la rue d'Ulm). Publication [74].

- Le machine learning informé par la physique combine la souplesse des approches basées sur les données avec l'interprétabilité des modèles physiques. Dans ce contexte, nous considérons un problème de régression où le risque empirique est régularisé par une équation aux dérivées partielles. Nous démontrons que pour des régularisations différentielles linéaires, le problème peut être formulé comme une tâche de régression à noyau. En exploitant alors la théorie des espaces de Hilbert à noyau reproduisant, nous obtenons des vitesses de convergence pour le minimiseur  $\hat{f}_n$  du risque régularisé, et montrons que  $\hat{f}_n$  converge au moins au taux minimax Sobolev. Cependant, des vitesses plus rapides peuvent être atteintes, ce qui montre que des informations de nature physique améliorent, le cas échéant, les performances statistiques des estimateurs. En collaboration avec Francis Bach (Directeur de recherche, Inria Paris), Claire Boyer et Nathan Doumèche. Publication [77].

## Bioinformatique

- Standardisation et analyse statistique des séquences de nucléotides de la base IMGT relatives aux jonctions TRAV-TRAJ des récepteurs de la cellule T. En collaboration avec Kevin Bleakley, Véronique Guidicelli (Ingénieur d'études, Université Montpellier 2), Marie-Paule Lefranc (Professeur, Université Montpellier 2) et Jan Wu (Chercheuse, Université Montpellier 2). Publications [25].
- Développement d'un nouvel algorithme de reconstruction de réseaux biologiques basés sur des modèles SVM (*Support Vector Machines*) locaux. Mise en évidence des excellentes propriétés de la méthode en termes de courbes ROC (*Receiver Operating Characteristic*) et FDR (*False Discovery Rate*) sur un réseau métabolique et un réseau d'interactions protéines-protéines. En collaboration avec Kevin Bleakley et Jean-Philippe Vert. Publication [27].
- Analyse du biais statistique des outils automatiques de la base IMGT relatifs au processus de grignotage des jonctions TRA et TRG des récepteurs de la cellule T. En collaboration avec Kevin Bleakley et Marie-Paule Lefranc. Publication [35].
- Dans ce travail didactique à destination de la communauté biostatistique et bioinformatique, nous passons en revue et comparons les performances théoriques et pratiques de quelques techniques essentielles en classification supervisée (plus proches voisins, forêts aléatoires, SVM, etc.), qu'il s'agisse du cadre binaire ou du cadre multiclasse. En collaboration avec une équipe internationale de sept chercheurs constituée à la suite du *Workshop on Probability Estimation in Prognostics* tenu à Zürich en septembre 2011. Publication [51].

## Congrès, workshops, séminaires et séjours de recherche \_\_\_\_\_

Je me suis exprimé au sein de 81 congrès nationaux ou internationaux (dont 69 en qualité d'invité) et lors de 83 colloquiums, séminaires ou groupes de travail. En voici le détail.

### Exposés oraux lors de congrès et workshops

- *7th International Meeting on Statistical Climatology*, Whistler, Canada (mai 1998).
- *GeoENV'98*, Valence, Espagne (novembre 1998).
- *XXXIèmes Journées de Statistique*, Grenoble, France (mai 1999).

- XXXIIèmes Journées de Statistique, Fès, Maroc (mai 2000).
- XXXIIIèmes Journées de Statistique, Nantes, France (mai 2001).
- XXXIVèmes Journées de Statistique, Bruxelles, Belgique (mai 2002).
- *International Conference on Current Advances and Trends in Nonparametric Statistics*, Hersonissos, Crète (juillet 2002, **invité**).
- *Statistical Learning, Theory and Applications*, CNAM, France (novembre 2002, **invité**).
- *Asymptotical Statistics of Stochastic Processes IV*, Le Mans, France (décembre 2002, **invité**).
- XXXVèmes Journées de Statistique, Lyon, France (juin 2003, **invité**).
- XXXVIèmes Journées de Statistique, Montpellier, France (mai 2004).
- *Journées Données Fonctionnelles*, Rennes, France (septembre 2004, **invité**).
- *Journée Optimisation et Statistiques*, Avignon, France (février 2005, **invité**).
- *Joint BeNeLuxFra Conference in Mathematics*, Gand, Belgique (mai 2005, **invité**).
- XXXVIIèmes Journées de Statistique, Pau, France (juin 2005, **invité**).
- *25th European Meeting of Statisticians*, Oslo, Norvège (juillet 2005, **invité**).
- *Rencontres de Statistiques Mathématiques 5*, CIRM de Luminy, France (décembre 2005).
- XXXVIIIèmes Journées de Statistique, Clamart, France (juin 2006).
- *Journées MAS de la SMAI – Modèles spatiaux*, Lille, France (septembre 2006, **invité**).
- *Journées de Statistiques, Rennes 2006*, Rennes, France (octobre 2006, **invité**).
- *Statistique Mathématique et Applications*, CIRM de Luminy, France (novembre 2006).
- XXXIXèmes Journées de Statistique, Angers, France (juin 2007).
- *GyörfiFest: Information Theory and Statistics Workshop in honor of Professor László Györfi's 60'th birthday*, Budapest, Hongrie (octobre 2007, **invité**).
- *Journée de Statistique*, Toulouse, France (décembre 2007, **invité**).
- *Workshop on Representations and Inference on Probability Distributions*, NIPS Conference, Whistler, Canada (décembre 2007, **invité**).
- *Mathematics for Biological Networks*, Paris, France (décembre 2007, **invité**).
- *Multimodality and Related Topics*, Nanterre, France (mars 2008, **invité**).
- *Congrès Conjoint de la Société Statistique du Canada et de la Société Française de Statistique*, Ottawa, Canada (mai 2008, **invité**).
- *STATDEP 2008: Statistics for Dependent Data*, Paris, France (juin 2008, **invité**).
- *IWAP 2008: International Workshop on Applied Probability*, Compiègne, France (juillet 2008, **invité**).

- *37th Annual Meeting of Dutch Statisticians and Probabilists*, Lunteren, Pays-Bas (novembre 2008, **invité**).
- *Workshop on Sparsity and Inverse Problems in Statistical Theory and Econometrics*, Berlin, Allemagne (décembre 2008, **invité**).
- *The 57th Session of the International Statistics Institute*, Durban, Afrique du Sud (août 2009, **invité**).
- *Journées STAR 2009*, Rennes, France (octobre 2009, **invité**).
- *42èmes Journées de Statistique*, Marseille, France (mai 2010).
- *IWAP 2010: International Workshop on Applied Probability*, Madrid, Espagne (juillet 2010, **invité**).
- *Prague Stochastics 2010*, Prague, Tchéquie (septembre 2010, **invité**).
- *Le Laboratoire de Probabilités et Modèles Aléatoires fête ses 50 ans !!*, Paris, France (décembre 2010, **invité**).
- *The Eighth ICSA International Conference: Frontiers of Interdisciplinary and Methodological Statistical Research*, Canton, Chine (décembre 2010, **invité**).
- *High Dimensional Statistics: Advances and Challenges*, Singapour (mai 2011, **invité**).
- *2011 Joint Statistical Meetings*, Miami Beach, Etats-Unis (août 2011, **invité**).
- *Workshop on Probability Estimation in Prognostics*, Zürich, Suisse (septembre 2011, **invité**).
- *Rencontres de Statistique Avignon-Marseille*, Marseille, France (juin 2012, **invité**).
- *Workshop on Modern Nonparametric Methods for Time Series, Reliability & Optimization*, Heverlee, Belgique (septembre 2012, **invité**).
- *Mathématiques et Grandes Dimensions: De la Théorie aux Développements Industriels*, Lyon, France (décembre 2012, **invité**).
- *Approximate Bayesian Computation in Rome*, Rome, Italie (mai 2013, **invité**).
- *Journée Mathématiques-Géophysique – Action Tarantola*, Paris, France (septembre 2013, **invité**).
- *Advances in Multivariate Analysis*, Lille, France (novembre 2013, **invité**).
- *Méthodes Mathématiques du Traitement d'Images*, Paris, France (décembre 2013, **invité**).
- *Horizon Maths 2013*, Paris, France (décembre 2013, **invité**).
- *The Ninth ICSA International Conference: Challenges of Statistical Methods for Interdisciplinary Research and Big Data*, Hong Kong, Chine (décembre 2013, **invité**).
- *XIII CLAPEM: Latin American Congress of Probability and Mathematical Statistics*, Carthagène des Indes, Colombie (septembre 2014, **invité, conférencier plénier**).
- *Journée de l'équipe Probabilités, Analyse et Statistique*, Clermont Ferrand, France (novembre 2014, **invité**).

- *Forum Franco-Uruguayen des Sciences, des Technologies et de l'Education pour un développement durable*, Montevideo, Uruguay (décembre 2014, **invité**).
- *ENAR 2015 Spring Meeting*, Miami, Etats-Unis (mars 2015, **invité**).
- *Methodological Advances in Statistics Related to Big Data*, Castro Urdiales, Espagne (juin 2015, **invité, conférencier plénier**).
- *Statistics Workshop*, Helsinki, Finlande (avril 2016, **invité**).
- *Networks: Learning, Information and Complexity*, Paris, France (mai 2016, **invité**).
- *Third Conference of the International Society for NonParametric Statistics*, Avignon, France (juin 2016, **invité, conférencier plénier**).
- *COMPSTAT 2016*, Oviedo, Espagne (août 2016, **invité, conférencier plénier**).
- *Statistical Learning and Econometrics*, Rotterdam, Pays-Bas (mai 2017, **invité**).
- *MASCOT-NUM 2018*, Nantes, France (mars 2018, **invité**).
- *New Developments in Statistics: Massive (Big) Data a Challenge or a Curse for Statistics?*, Heverlee, Belgique (juillet 2018, **invité**).
- *Structural Inference in High-Dimensional Models*, Moscou, Russie (septembre 2018, **invité**).
- *Optimization and Learning*, Toulouse, France (septembre, **invité**).
- *Journées des Prix Mathématiques 2018 de l'Académie des Sciences*, Rennes, France (décembre 2018, **invité**).
- *Montpellier, berceau de la data science. Colloque en l'honneur du Pr Yves Escoufier*, Montpellier, France (juin 2019, **invité**).
- *Indo-French Joint Workshop on Statistics and Artificial Intelligence for Data Science*, Kolkata, Inde (janvier 2020, **invité**).
- *French-German Machine Learning Symposium*, Munich, Allemagne (virtuel, mai 2021, **invité**).
- *5th International Workshop on Functional and Operatorial Statistics*, Brno, République Tchèque (virtuel, juin 2021, **invité, conférencier plénier**).
- *Journée du GdR ISIS sur la Théorie du Deep Learning* (virtuel, juin 2021, **invité, conférencier plénier**).
- *Emeritus Celebration Jan Beirlant*, Leuven, Belgique (septembre 2021, **invité**).
- *Journée de la Fédération Bourgogne Franche-Comté Mathématiques*, Dijon, France (novembre 2021, **invité**).
- *Workshop on Multivariate Analysis of Complex Data*, Bruxelles, Belgique (mai 2022, **invité**).
- *Journées MAS 2022*, Rouen, France (août 2022, **invité, conférencier plénier**).
- *2022 IMS International Conference on Statistics and Data Science*, Florence, Italie (décembre 2022, **invité**).

- *Winter School on Deep Learning 2023*, Kolkata, Inde (virtuel, janvier 2023, **invité, conférencier plénier**).
- *IWAP 2023: 10th International Workshop on Applied Probability*, Thessalonique, Grèce (juin 2023, **invité, conférencier plénier**).
- *European Meeting of Statisticians 2023*, Varsovie, Pologne (juillet 2023, **invité, conférencier plénier**, deux *Forum Lectures*).
- *XVI<sup>e</sup> Journées de Géostatistique*, Fontainebleau, France (septembre 2023, **invité**).
- *Statistics and Data Science 2024*, Palerme, Italie (avril 2024, **invité, conférencier plénier**).

### Exposés oraux lors de séminaires et groupes de travail

- Séminaire de Probabilités et Statistique, Université Montpellier 2 (mai 1998, mai 1999, mars 2000, juin 2002, octobre 2004).
- Séminaire de Statistique et Econométrie, Université Toulouse 1 (octobre 2000, décembre 2004).
- Séminaire de Statistique, Université Toulouse III (novembre 2000, février 2005, janvier 2015).
- Séminaire de Statistique, Université de Montréal (février 2001).
- Séminaire de Statistique et Modélisation Stochastique, Université Grenoble I (avril 2001).
- Séminaire de Statistique Non-paramétrique et Semi-paramétrique, Laboratoire de Statistique Théorique et Appliquée, Université Paris VI (octobre 2001, juin 2003).
- Groupe de travail de Statistique, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris VI et Paris Diderot – Paris 7 (janvier 2002, mars 2005, janvier 2007, février 2009).
- Séminaire Parisien de Statistique, Institut Henri Poincaré (avril 2002, novembre 2002, septembre 2009).
- Groupe de travail Probabilités Numériques, Statistique des Processus et Finance, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris VI et Paris Diderot – Paris 7 (novembre 2002).
- Séminaire d'Econométrie et de Statistique, Université Libre de Bruxelles (décembre 2002, octobre 2008, septembre 2016).
- Séminaire de Probabilités et Statistique, Université de Provence – Aix-Marseille 1 (mars 2003).
- Groupe de travail de Statistiques Spatiales, Institut National Agronomique Paris-Grignon (juillet 2003).
- Séminaire de Statistique, Université Rennes 2 (février 2004).
- Séminaire Probabilités et Statistique, Université Paris-Sud (mars 2004, octobre 2009, septembre 2018).



- Groupe de travail Support Vector Machines et Estimation Adaptative, Ecole Normale Supérieure de la rue d’Ulm (avril 2004).
- Séminaire de Probabilités et Statistiques, Université de Versailles Saint-Quentin-en-Yvelines (mai 2004).
- Séminaire de Probabilités et Statistiques, Université Nice Sophia Antipolis (novembre 2005, novembre 2010).
- Séminaire de Statistique, Université de Rennes 1 (janvier 2006).
- Conférence d’Initiation à la Recherche, Ecole Normale Supérieure de Rennes (janvier 2006, février 2014).
- Séminaire de Statistique, Université Libre de Bruxelles (mars 2006).
- Statistics and Operational Research Seminar, Université Pompeu Fabra de Barcelone (novembre 2006).
- Séminaire Applications des Mathématiques, Université de Bourgogne (décembre 2006).
- Séminaire de Probabilités, Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris VI et Paris Diderot – Paris 7 (octobre 2007).
- Séminaire de Mathématiques Appliquées, Université Bordeaux 1 (janvier 2008).
- Groupe de travail Apprentissage, Ecole Normale Supérieure de la rue d’Ulm (février 2008).
- Groupe de travail Applications des Mathématiques, Ecole Normale Supérieure de Rennes (avril 2008).
- Séminaire du Département OSIRIS, EDF Recherche & Développement, Clamart (octobre 2008).
- Groupe de travail en Biostatistiques, Université Henri Poincaré – Nancy 1 (octobre 2008).
- Groupe de travail en Statistiques, Université Paris Descartes (février 2009).
- Séminaire Business Intelligence, Télécom ParisTech (avril 2009).
- Séminaire de Statistique, AgroParisTech (octobre 2009).
- Séminaire de Probabilité-Statistique, Université de Franche-Comté (janvier 2010).
- Séminaire de Probabilités et Statistiques, Université Lille 1 (mai 2010).
- Séminaire de Statistique, Université Toulouse 1 (octobre 2010, mars 2015).
- Colloquium du MAP5, Université Paris Descartes (février 2011).
- Séminaire de Mathématiques Appliquées, Université Blaise Pascal (février 2011).
- Seminar Series in Finance, Cyprus University of Technology (mars 2011).
- Groupe de travail Prévision, ENGREF (mai 2011).
- Groupe de travail Analyse, Probabilités et Statistique, Université Paris-Est Marne-la-Vallée (mai 2011).

- Séminaire Modèles Aléatoires et Applications, Université de Caen Basse-Normandie (septembre 2011).
- Séminaire MODAL'X, Université Paris Ouest – Nanterre La Défense (novembre 2011).
- Statistical Laboratory Seminar, Université de Cambridge (janvier 2012).
- Séminaire SMILE in Paris, Ecole des Mines de Paris (mars 2012).
- Séminaire Point de vue, Université Paris Diderot – Paris 7 (novembre 2012).
- Séminaire Maths-Info, Université Paris VI (novembre 2013).
- Séminaire invité, Université de Liège (octobre 2014).
- Colloquium Lorrain de Mathématiques, Université de Lorraine (octobre 2014).
- Machine Learning External Seminar, University College London (janvier 2015).
- Séminaire de Statistique, Université de Saint-Jacques-de-Compostelle (mars 2015).
- Joint Artificial Intelligence & Statistics Seminar, Université Cornell (mars 2015).
- Statistics, Applied Probability & OR Seminar, Université d'Oxford (octobre 2015).
- Séminaire du Yu Research Group, Université de Californie à Berkeley (février 2018).
- Probability & Statistics Seminar, Université du Luxembourg (virtuel, octobre 2020, février 2024).
- Research Seminars Series in Financial Engineering and Operations Research, Université Mohammed VI Polytechnique (novembre 2021).
- Séminaire DATAIA, Institut DATAIA (janvier 2022).
- Abu Dhabi Stochastics Seminar, New York University Abu Dhabi et Sorbonne University Abu Dhabi (virtuel, mars 2022).
- Séminaire Palaisien, ENSAE Paris (avril 2022).
- Mathematics Colloquium, Université de Coimbra (juillet 2022).
- Statistics Department Seminar, Université de Floride (virtuel, octobre 2022).
- Statistics Seminars, Université Bocconi (mai 2023).
- Colloquium de l'Institut de Mathématiques de Jussieu-Paris Rive Gauche (décembre 2023).
- Colloquium Primavera 2024, Université Autonome de Madrid (mars 2024).

## Organisation de colloques et sessions

- Coorganisateur avec Gilbert Saporta (Professeur, Conservatoire National des Arts et Métiers) de la session invitée *Functional Data Analysis* au sein de l'*International Symposium on Applied Stochastic Models and Data Analysis*, Brest, France (mai 2005).
- Membre du comité d'organisation des *Premières Rencontres des Jeunes Statisticiens*, Aussois, France (septembre 2005).

- Coorganisateur avec Yannick Baraud (Professeur, Université du Luxembourg), Laurent Cavalier (Professeur, Université de Provence – Aix-Marseille 1), Susan Holmes (Professeur, Université Stanford) et Béatrice Laurent-Bonneau (Professeur, INSA Toulouse) des *1ères Journées Statistiques du Sud*, Nice, France (avril 2007).
- Membre du comité de programme des *XXXIXèmes Journées de Statistique*, Angers, France (juin 2007).
- Coorganisateur avec Bernard Garel (Professeur, Institut National Polytechnique de Toulouse) des *Deuxièmes Rencontres des Jeunes Statisticiens*, Aussois, France (septembre 2007).
- Coorganisateur avec Yannick Baraud, Laurent Cavalier, Béatrice Laurent-Bonneau, Jean-Michel Loubes (Professeur, Université Toulouse III) et Gábor Lugosi des *2èmes Journées Statistiques du Sud*, Toulouse, France (juin 2008).
- Président du comité de programme des *41èmes Journées de Statistique*, Bordeaux, France (mai 2009).
- Coorganisateur avec Yannick Baraud, Laurent Cavalier, Béatrice Laurent-Bonneau, Jean-Michel Loubes, Gábor Lugosi et Jean-Michel Marin (Professeur, Université Montpellier 2) des *3èmes Journées Statistiques du Sud*, Porquerolles, France (juin 2009).
- Membre du comité d'organisation de *COMPSTAT'2010*, Paris, France (août 2010).
- Coorganisateur avec Yannick Baraud, Laurent Cavalier, Béatrice Laurent-Bonneau, Jean-Michel Loubes, Gábor Lugosi et Jean-Michel Marin des *5èmes Journées Statistiques du Sud*, Nice, France (juin 2011).
- Coorganisateur avec Pierre Alquier de la session invitée *High-Dimensional Statistics, Sparsity and Applications* au sein de la *4th International Conference of the ERCIM Working Group on Computing & Statistics*, Londres, Royaume-Uni (décembre 2011).
- Membre du comité de programme de *EGC 2012*, Bordeaux, France (février 2012).
- Coorganisateur avec Yannick Baraud, Laurent Cavalier, Béatrice Laurent-Bonneau, Jean-Michel Loubes, Gábor Lugosi et Jean-Michel Marin des *6èmes Journées Statistiques du Sud*, Toulouse, France (juin 2012).
- Membre du comité d'organisation du colloque *Unsupervised Learning and High-Dimensional Statistics*, Paris, France (septembre 2013).
- Membre du comité d'organisation du colloque *Horizons de la Statistique*, Paris, France (janvier 2014).
- Coorganisateur avec Yannick Baraud, Laurent Cavalier, Béatrice Laurent-Bonneau, Jean-Michel Loubes, Gábor Lugosi et Jean-Michel Marin des *7èmes Journées Statistiques du Sud*, Barcelone, Espagne (juin 2014).
- Président du comité d'organisation du colloque *Statistique et Données Massives : Enjeux et Perspectives*, Paris, France (janvier 2015).
- Membre du comité d'organisation du colloque *Horizon Maths 2015 : Santé et Données*, Paris, France (décembre 2015).
- Membre du comité scientifique de la semaine thématique *Apprentissage*, CIRM de Luminy, France (février 2016).

- Membre du comité de programme scientifique du *European Meeting of Statisticians*, Helsinki, Finlande (juillet 2017).
- Coorganisateur avec Giles Hooker (Professeur, Université Cornell), Lucas Mentch (Assistant professor, Université de Pittsburgh) et Stefan Wager (Assistant professor, Université Stanford) du *Workshop on the Interface of Machine Learning and Statistical Inference*, Banff International Research Station, Canada (janvier 2018).
- Coorganisateur avec Gilles Blanchard (Professeur, Université Paris-Saclay), Stéphane Mallat (Professeur, Collège de France), Pascal Massart (Professeur, Université Paris-Saclay) et Dominique Picard (Professeur émérite, Université Paris Cité) de la conférence *Math & IA*, Paris, France (mars 2021).
- Coorganisateur de la conférence *NeurIPS@Paris 2021-22-23*, Paris, France (décembre 2021, novembre 2022 et décembre 2023).
- Membre du comité de programme de la *2022 IMS International Conference on Statistics and Data Science*, Florence, Italie (décembre 2022).

### Séjours de recherche à l'étranger

- Séjour de recherche de 6 mois au Helmholtz-Zentrum Geesthacht (Allemagne, janvier à juin 1997), laboratoire de Hans von Storch.
- Chercheur invité à l'Université Catholique de Leuven (Belgique, février 2000 et juillet 2001), laboratoire de Jan Beirlant.
- Séjour postdoctoral à l'Université McGill de Montréal (Canada, février 2001), laboratoire de Luc Devroye.
- Chercheur invité par Luc Devroye à l'Université Catholique de Louvain (Belgique, mai 2002) et à l'Université McGill de Montréal (Canada, février 2003, mars 2006, mars 2008, février 2010, février 2011, février 2012, février 2013, août 2013, février 2014, août 2014, février 2015, février 2016, février 2017 et mars 2019).
- Séjour à l'Université Florida State de Tallahassee (Etats-Unis, juillet 2003), laboratoire de Florentina Bunea et Marten Wegkamp.
- Chercheur invité à l'Université Pompeu Fabra de Barcelone (Espagne, janvier 2004 et octobre 2006), laboratoire de Gábor Lugosi.
- Chercheur invité à l'University of Technology and Economics de Budapest (Hongrie, février 2007), laboratoire de László Györfi.
- Chercheur invité à la Cyprus University of Technology de Limassol (Chypre, mars 2011), laboratoire de Yannis Yatracos.
- Séjour à l'Université Stanford (Etats-Unis, février 2018), laboratoire de Stefan Wager.

### Activités d'encadrement doctoral

---

Je bénéficie depuis le 1<sup>er</sup> octobre 2003 de la prime d'encadrement doctoral et de recherche (PEDR puis PIR).

### L3, DEA et M2

- Encadrement du stage de DEA de Matthieu Stasinski (DEA de statistique, Université Paris VI) sur le thème *Classification non paramétrique de courbes* (2003).
- Encadrement du stage de DEA de Idar Abarro (DEA de statistique, Université Paris VI) sur le thème *Méthodes combinatoires : théorie et pratique* (2004).
- Encadrement du stage de DEA de Kevin Bleakley (DEA de biostatistique, Université Montpellier 2) sur le thème *Apprentissage et reconstruction de graphes* (2005).
- Encadrement du stage de M2 de Thomas Laloë (M2 recherche, spécialité biostatistique, Université Montpellier 2) sur le thème *Régression et apprentissage statistique* (2006).
- Co-encadrement avec Franck Nicoud (Professeur, Université Montpellier 2) du stage de M2 de Jean-Joël Lalanne (M2 recherche, spécialité mathématiques, Université Montpellier 2) sur le thème *Etude des propriétés statistiques d'un flot turbulent* (2007).
- Co-encadrement avec Joannès Vermorel (fondateur de la société Lokad) du stage de M2 de Benoît Patra (M2 recherche, spécialité probabilités et applications, Université Paris VI) sur le thème *Prévision non paramétrique de séries temporelles* (2008).
- Co-encadrement avec Pascal Massart du stage de M2 de Clément Levrard (Ecole Normale Supérieure de la rue d'Ulm) sur le thème *Une méthode d'égalisation en sélection de modèle* (2010).
- Co-encadrement avec Romain Laroche (société Orange France) du stage de M2 de Rémi Lemmonier (M2 recherche, spécialité probabilités et modèles aléatoires, Université Paris VI) sur le thème *Construction d'un système de dialogue incrémental et recherche de la stratégie optimale* (2012).
- Encadrement du stage de M2 de Ryad Zenine (M2 recherche, spécialité probabilités et modèles aléatoires, Université Paris VI) sur le thème *Algorithmes en ligne distribués et asynchrones* (2014).
- Co-encadrement avec Yvon Maday (Professeur, Sorbonne Université), Josselin Garnier (Professeur, Ecole Polytechnique) et Erwan Scornet du stage de L3 d'Arthur Pajot (L3, Université Paris VI) dans le cadre d'une collaboration industrielle avec la société KILA Systems (2014).
- Encadrement du stage de M2 de Jaime Roquero Gimenez (Ecole Normale Supérieure de la rue d'Ulm) sur le thème *Affaiblissement d'algorithmes par relaxation convexe* (2015).
- Co-encadrement avec Benoît Cadre du stage de M2 d'Adeline Fermanian (ETH Zurich) sur le thème *Signature et apprentissage statistique* (2018).
- Co-encadrement avec Claire Boyer du stage de M2 de Linus Bleistein (Ecole Normale Supérieure de la rue d'Ulm) sur le thème *SiGANtures: Generating time series using Wasserstein-Generative Adversarial Nets and the signature transform* (2020).
- Encadrement du stage de M2 d'Arthur Stéphanovitch (M2 MVA, ENS Paris-Saclay) sur le thème *Propriétés mathématiques et statistiques des GANs* (2021).
- Encadrement du stage de M2 de Nathan Doumèche (Ecole Normale Supérieure de la rue d'Ulm) sur le thème *Développement de Taylor signé et applications à l'apprentissage statistique* (2021).

- Co-encadrement avec Pierre Marion du stage de M2 de Yu-Han Wu (Ecole Normale Supérieure de la rue d’Ulm) sur le thème *Deep learning theory* (2023).
- Co-encadrement avec Claire Boyer du stage de M2 de Guillhem Artis (M2 Statistique & Learning, Sorbonne Université) sur le thème *Towards a non-asymptotic theory of PINNs* (2024).
- Co-encadrement avec Claire Boyer et Pierre Marion du stage de quatrième année de Yu-Han Wu sur le thème *A mathematical analysis of diffusion models for generative modeling* (2024).

## Thèses

- Co-encadrement avec Alain Berlinet de la thèse de Laurent Rouvière (Maître de conférences, Université Rennes 2) sur le thème *Estimation de densité en dimension élevée et classification de courbes* (thèse commencée le 1<sup>er</sup> octobre 2002 et soutenue le 18 novembre 2005).
- Encadrement de la thèse de Kevin Bleakley (Chargé de recherche, Inria Saclay – Ile-de-France) sur le thème *Quelques contributions à l’analyse statistique et à la classification des graphes et des courbes. Applications à l’immunobiologie et à la reconstruction des réseaux biologiques* (thèse commencée le 1<sup>er</sup> octobre 2005 et soutenue le 23 novembre 2007).
- Co-encadrement avec Benoît Cadre de la thèse de Thomas Laloë (Maître de conférences, Université Côte d’Azur) sur le thème *Sur quelques problèmes d’apprentissage supervisé et non supervisé* (thèse commencée le 1<sup>er</sup> octobre 2006 et soutenue le 27 novembre 2009).
- Co-encadrement avec Bertrand Iooss (Ingénieur de recherche, EDF) de la thèse de Benjamin Auder (Ingénieur d’études CNRS, Université Paris-Saclay) sur le thème *Classification et modélisation de sorties fonctionnelles de codes de calcul. Application aux calculs thermo-hydrauliques accidentels dans les réacteurs à eau pressurisée (REP)* (thèse commencée le 1<sup>er</sup> janvier 2008 et soutenue le 5 mai 2011).
- Encadrement de la thèse d’Aurélié Fischer (Maître de conférences, Université Paris Cité) sur le thème *Apprentissage statistique non supervisé : grande dimension et courbes principales* (thèse commencée le 1<sup>er</sup> octobre 2008 et soutenue le 9 juin 2011).
- Encadrement de la thèse de Benoît Patra (chief technology officer, société iAdvize) sur le thème *Apprentissage à “grande échelle” : contribution à l’étude d’algorithmes de clustering répartis asynchrones* (thèse commencée le 1<sup>er</sup> décembre 2008 et soutenue le 6 mars 2012).
- Co-encadrement avec Pascal Massart de la thèse de Clément Levrard (Professeur, Université de Rennes) sur le thème *Quantification vectorielle en grande dimension : vitesses de convergence et sélection de variables* (thèse commencée le 1<sup>er</sup> septembre 2010 et soutenue le 30 septembre 2014).
- Co-encadrement avec Eric Moulines (Professeur, Ecole Polytechnique) de la thèse de Benjamin Guedj (Chargé de recherche, Inria Lille – Nord Europe) sur le thème *Agrégation d’estimateurs et de classificateurs : théorie et méthodes* (thèse commencée le 1<sup>er</sup> février 2011 et soutenue le 4 décembre 2013).
- Co-encadrement avec Olivier Lopez (Professeur, Sorbonne Université) de la thèse de Svetlana Gribkova (Maître de conférences, Université Paris Cité) sur le thème *Contribution à l’inférence statistique en présence de censure multivariée* (thèse commencée le 1<sup>er</sup> octobre 2011 et soutenue le 29 septembre 2014).

- Co-encadrement avec Bertrand Michel (Professeur, Ecole Centrale de Nantes) et Philippe Saint Pierre (Maître de conférences, Université Toulouse III) de la thèse de Baptiste Gregorutti (Ingénieur de recherche contractuel, Sorbonne Université) sur le thème *Forêts aléatoires et sélection de variables : analyse des données des enregistreurs de vol pour la sécurité aérienne* (thèse commencée le 1<sup>er</sup> décembre 2011 et soutenue le 11 mars 2015).
- Co-encadrement avec Jean-Philippe Vert de la thèse d’Erwan Scornet (Maître de conférences, Ecole Polytechnique) sur le thème *Apprentissage et forêts aléatoires* (thèse commencée le 1<sup>er</sup> septembre 2012 et soutenue le 29 novembre 2015).
- Co-encadrement avec Jean-Patrick Baudry (Maître de conférences, Sorbonne Université) et Laurent Rouvière de la thèse de Nedjmeddine Allab (data scientist, freelance) sur le thème *Détection d’anomalies et de ruptures dans les séries temporelles. Applications à la gestion de production de l’électricité* (thèse commencée le 1<sup>er</sup> décembre 2012 et soutenue le 21 novembre 2016).
- Co-encadrement avec Sylvain Arlot (Professeur, Université Paris-Saclay) de la thèse de Damien Garreau (Maître de conférences, Université Côte d’Azur) sur le thème *Détection de ruptures et méthodes à noyaux* (thèse commencée le 1<sup>er</sup> septembre 2014 et soutenue le 12 octobre 2017).
- Co-encadrement avec Bertrand Michel et Philippe Saint Pierre de la thèse de Nazih Benoumechiara (quant, société Qube Research & Technologies) sur le thème *Traitement de la dépendance en analyse de sensibilité pour la fiabilité industrielle* (thèse commencée le 1<sup>er</sup> juillet 2015 et soutenue le 10 juillet 2019).
- Co-encadrement avec Pierre Bougnères (Professeur-Praticien Hospitalier, Université Paris-Sud 11) de la thèse de Félix Balazard (director of optimized development, société Owkin) sur le thème *Contributions à la génétique et l’épidémiologie des maladies complexes pour une médecine personnalisée* (thèse commencée le 1<sup>er</sup> septembre 2015 et soutenue le 17 décembre 2018).
- Co-encadrement avec Maxime Sangnier de la thèse d’Ugo Tanielian (senior AI engineer, société Gleamer) sur le thème *“Generative Adversarial Networks”: théorie et pratique* (thèse commencée le 1<sup>er</sup> février 2018 et soutenue le 23 avril 2021).
- Co-encadrement avec Bertrand Michel de la thèse de Panayotis Papoutsis (Data Scientist, société Finclude) sur le thème *Potentiel et prévision des temps d’attente pour le covoiturage sur un territoire* (thèse commencée le 1<sup>er</sup> mars 2018 et soutenue le 17 décembre 2021).
- Co-encadrement avec Erwan Scornet de la thèse de Clément Bénard (Ingénieur de recherche, société Safran) sur le thème *Forêts aléatoires et interprétabilité des algorithmes d’apprentissage* (thèse commencée le 1<sup>er</sup> novembre 2018 et soutenue le 6 décembre 2021).
- Co-encadrement avec Patrick Gallinari (Professeur, Sorbonne Université) de la thèse de Jérémie Donà (Data Scientist, société InstaDeep) sur le thème *Apprentissage statistique de dynamiques physiques* (thèse commencée le 1<sup>er</sup> novembre 2018 et soutenue le 25 mai 2022).
- Co-encadrement avec Benoît Cadre de la thèse d’Adeline Fermanian (chercheuse en machine learning, société Califrais) sur le thème *Learning time-dependent data with the signature transform* (thèse commencée le 1<sup>er</sup> novembre 2018 et soutenue le 25 octobre 2021).
- Co-encadrement avec Jean-Philippe Vert de la thèse de Pierre Marion (postdoctorant, EPFL) sur le thème *Mathematics of deep learning: generalization, optimization, continuous-time models* (thèse commencée le 1<sup>er</sup> octobre 2020 et soutenue le 20 novembre 2023).

- Co-encadrement avec Claire Boyer de la thèse de Nathan Doumèche (Sorbonne Université et EDF) sur le thème *Physics-Informed Machine Learning pour la prévision en périodes atypiques : application à la prévision de consommation d'électricité française et à la prévision de production d'énergie renouvelable* (thèse commencée le 1<sup>er</sup> janvier 2023).
- Co-encadrement avec Laurent Rouvière de la thèse d'Alexandre Borel (Sorbonne Université et Crédit Agricole Corporate & Investment Bank) sur le thème *Méthodes quantitatives et machine learning en finance de marché : extraction d'information et de relations dans un contexte de prédiction* (thèse commencée le 1<sup>er</sup> mars 2023).

## Postdocs

- Co-encadrement avec Raphaël Porcher du postdoctorat de Qiming Du (research associate, société Capital Fund Management) dans le cadre du projet ANR OPT (*Optimization of Personalized Treatments*, 2020-2021).

## Participations à des jurys

---

### Thèses

- Membre du jury de la thèse de Laurent Rouvière, Université Montpellier 2 (novembre 2005).
- Rapporteur de la thèse de Christine Tuleau, Université Paris-Sud 11 (décembre 2005).
- Membre du jury de la thèse de Nicolas Klutchnikoff, Université de Provence – Aix-Marseille 1 (décembre 2005).
- Rapporteur de la thèse de Tristan Mary-Huard, Université Paris-Sud 11 (juillet 2006).
- Rapporteur de la thèse de Baba Thiam, Université de Versailles Saint-Quentin-en-Yvelines (décembre 2006).
- Président du jury de la thèse d'Ahmad Younso, Université Montpellier 2 (avril 2007).
- Rapporteur de la thèse de Vincent Lefieux, Université Rennes 2 (octobre 2007).
- Membre du jury de la thèse de Kevin Bleakley, Université Montpellier 2 (novembre 2007).
- Membre du jury de la thèse de Roxanne Jallet, Université Paris VI (janvier 2008).
- Rapporteur de la thèse de Bertrand Michel, Université Paris-Sud 11 (septembre 2008).
- Président du jury de la thèse de Franck Rapaport, Université Paris VI (septembre 2008).
- Membre du jury de la thèse de Kevin Loquin, Université Montpellier 2 (novembre 2008).
- Membre du jury de la thèse de Zaïd Harchaoui, Télécom ParisTech (novembre 2008).
- Rapporteur de la thèse de Sébastien Loustau, Université de Provence – Aix-Marseille 1 (novembre 2008).
- Membre du jury de la thèse de Frédéric Guilloux, Université Paris Diderot – Paris 7 (décembre 2008).
- Membre du jury de la thèse de Boris Labrador, Université Paris VI (décembre 2008).
- Rapporteur de la thèse d'Alain Celisse, Université Paris-Sud 11 (décembre 2008).



- Membre du jury de la thèse de Matthieu Cornec, Université Paris Ouest – Nanterre La Défense (juin 2009).
- Président du jury de la thèse de Julien Cornebise, Université Paris VI (juin 2009).
- Membre du jury de la thèse de Lahcen Douge, Université Paris VI (octobre 2009).
- Président du jury de la thèse d'Olivier Bouaziz, Université Paris VI (novembre 2009).
- Membre du jury de la thèse de Sophie Dede, Université Paris VI (novembre 2009).
- Membre du jury de la thèse de Thomas Laloë, Université Montpellier 2 (novembre 2009).
- Membre du jury de la thèse de Claire Coiffard, Université Paris VI (novembre 2009).
- Rapporteur de la thèse de Jean-Patrick Baudry, Université Paris-Sud 11 (décembre 2009).
- Membre du jury de la thèse d'Issam Elhattab, Université Paris VI (janvier 2010).
- Membre du jury de la thèse de Rémi Servien, Université Montpellier 2 (mars 2010).
- Rapporteur de la thèse de Nathanaël Mayo, Université Paris 1 Panthéon-Sorbonne (mars 2010).
- Membre du jury de la thèse de Tarek Zari, Université Paris VI (mai 2010).
- Président du jury de la thèse de Vianney Perchet, Université Paris VI (juin 2010).
- Président du jury de la thèse de Camille Sabbah, Université Paris VI (juillet 2010).
- Président du jury de la thèse d'Anne Cori, Université Paris VI (octobre 2010).
- Membre du jury de la thèse de Mohamed Cherfi, Université Paris VI (novembre 2010).
- Rapporteur de la thèse de Robin Genuer, Université Paris-Sud 11 (novembre 2010).
- Président du jury de la thèse de Thi Mong Ngoc Nguyen, Université Bordeaux 1 (novembre 2010).
- Membre du jury de la thèse de José Francisco Zelasco, Université Montpellier 2 (décembre 2010).
- Membre du jury de la thèse de Benjamin Auder, Université Paris VI (mai 2011).
- Membre du jury de la thèse d'Aurélien Fischer, Université Paris VI (juin 2011).
- Membre du jury de la thèse de Paul Rochet, Université Toulouse III (décembre 2011).
- Membre du jury de la thèse d'Elena Di Bernardino, Université Claude Bernard Lyon 1 (décembre 2011).
- Membre du jury de la thèse de Benoît Patra, Université Paris VI (mars 2011).
- Membre du jury de la thèse de Mohamed Achibi, Université Paris VI (juillet 2012).
- Membre du jury de la thèse de Moïse Jérémie, Université Paris VI (septembre 2012).
- Membre du jury de la thèse de Virgile Caron, Université Paris VI (octobre 2012).
- Membre du jury de la thèse de Caroline Meynet, Université Paris-Sud 11 (novembre 2012).

- Rapporteur de la thèse d'Ekaterina Sergienko, Université Toulouse III (novembre 2012).
- Président du jury de la thèse de Nicolas Jégou, Université Rennes 2 (novembre 2012).
- Rapporteur de la thèse de Christophe Denis, Université Paris Descartes (novembre 2012).
- Président du jury de la thèse d'Emmanuel Onzon, Université Paris VI (novembre 2012).
- Membre du jury de la thèse de Sarah Ouadah, Université Paris VI (décembre 2012).
- Président du jury de la thèse de Sylvain Girard, Ecole des Mines de Paris (décembre 2012).
- Membre du jury de la thèse de Ludovic Trinquart, Université Paris Descartes (mars 2013).
- Rapporteur de la thèse de Raphaël Coudret, Université Bordeaux 1 (septembre 2013).
- Président du jury de la thèse de Yannick Spill, Institut Pasteur (septembre 2013).
- Rapporteur de la thèse de Cindie Andrieu, Université Toulouse III (septembre 2013).
- Membre du jury de la thèse de Quentin Paris, Ecole Normale Supérieure de Rennes (octobre 2013).
- Président du jury de la thèse de Matthieu Solnon, Université Paris VI (novembre 2013).
- Membre du jury de la thèse de Benjamin Guedj, Université Paris VI (décembre 2013).
- Président du jury de la thèse d'Antoine Channarond, AgroParisTech (décembre 2013).
- Président du jury de la thèse de Loyal El Hajj, Université Paris VI (décembre 2013).
- Président du jury de la thèse de Saïp Ciss, Université Paris Ouest – Nanterre La Défense (juin 2014).
- Président du jury de la thèse de Sultana Didi, Université Paris VI (septembre 2014).
- Membre du jury de la thèse de Svetlana Gribkova, Université Paris VI (septembre 2014).
- Membre du jury de la thèse de Clément Levrard, Université Paris-Sud 11 (septembre 2014).
- Membre du jury de la thèse de Przemyslaw Sloma, Université Paris VI (septembre 2014).
- Membre du jury de la thèse d'Emilie Kaufmann, Télécom ParisTech (octobre 2014).
- Membre du jury de la thèse de Gilles Louppe, Université de Liège (octobre 2014).
- Membre du jury de la thèse de Khalid Chokri, Université Paris VI (novembre 2014).
- Président du jury de la thèse de Thomas Mainguy, Ecole Normale Supérieure de la rue d'Ulm (décembre 2014).
- Rapporteur de la thèse de Leslie Hatton, Agrocampus Ouest (janvier 2015).
- Président du jury de la thèse d'Alexis Decurninge, Université Paris VI (janvier 2015).
- Rapporteur de la thèse de Paula Saavedra Nieves, Université de Saint-Jacques-de-Compostelle (mars 2015).
- Membre du jury de la thèse de Baptiste Gregorutti, Université Paris VI (mars 2015).

- Membre du jury de la thèse d’Ilaria Giuliani, Ecole Normale Supérieure de la rue d’Ulm (septembre 2015).
- Membre du jury de la thèse d’Erwan Scornet, Université Paris VI (novembre 2015).
- Président du jury de la thèse d’Andrea Resmini, Université Paris VI (décembre 2015).
- Rapporteur de la thèse de José Luis Torrecilla Nogueras, Université Autonome de Madrid (décembre 2015).
- Membre du jury de la thèse de Meitner Cadena, Université Paris VI (janvier 2016).
- Président du jury de la thèse de Boris Béranger, Université Paris VI (janvier 2016).
- Membre du jury de la thèse de Matthieu Clertant, Université Paris VI (juin 2016).
- Membre du jury de la thèse d’ElMokhtar EzZahdi Alaya, Université Paris VI (juin 2016).
- Membre du jury de la thèse d’Assia Boumahdaf, Université Paris VI (septembre 2016).
- Président du jury de la thèse de Joon Kwon, Université Paris VI (octobre 2016).
- Président du jury de la thèse d’Arnaud de Myttenaere, Université Paris 1 Panthéon-Sorbonne (novembre 2016).
- Membre du jury de la thèse de Nedjmeddine Allab, Université Paris VI (novembre 2016).
- Président du jury de la thèse de Le Quyen Thieu, Université Paris VI (novembre 2016).
- Président du jury de la thèse de Nicolas Goix, Télécom ParisTech (novembre 2016).
- Membre du jury de la thèse d’Olivier Sorba, Université Paris-Sud (février 2017).
- Rapporteur de la thèse d’Hiba Alawieh, Université de Lille 1 (mars 2017).
- Président du jury de la thèse de The Tien Mai, ENSAE ParisTech (juin 2017).
- Rapporteur de la thèse de Thomas Berrett, Université de Cambridge (septembre 2017).
- Membre du jury de la thèse de Damien Garreau, Ecole Normale Supérieure de la rue d’Ulm (octobre 2017).
- Président du jury de la thèse de Guillaume Papa, Télécom ParisTech (janvier 2018).
- Rapporteur de la thèse de Sami Capderou, Université de Bordeaux (septembre 2018).
- Président du jury de la thèse d’Elsa Cazelles, Université de Bordeaux (septembre 2018).
- Rapporteur de la thèse de Claire Brécheteau, Université Paris-Sud (septembre 2018).
- Président du jury de la thèse d’Audrey Poterie, INSA Rennes (octobre 2018).
- Président du jury de la thèse de Guillermo Durand, Sorbonne Université (novembre 2018).
- Président du jury de la thèse de Le Li, Université d’Angers (novembre 2018).
- Président du jury de la thèse de Maxime Beauchamp, Mines ParisTech (décembre 2018).
- Membre du jury de la thèse de Félix Balazard, Sorbonne Université (décembre 2018).

- Président du jury de la thèse de Roy El Hourany, Sorbonne Université (mars 2019).
- Président du jury de la thèse de Diala Wehbe, Université de Lille (avril 2019).
- Membre du jury de la thèse de Geneviève Robin, Ecole Polytechnique (juin 2019).
- Président du jury de la thèse de Yohann Le Faou, Sorbonne Université (octobre 2019).
- Membre du jury de la thèse de Yating Liu, Sorbonne Université (décembre 2019).
- Membre du jury de la thèse de Benjamin Goehry, Université Paris-Saclay (décembre 2019).
- Président du jury de la thèse d'Arthur Pajot, Sorbonne Université (décembre 2019).
- Membre du jury de la thèse de Maximilien Baudry, Université de Lyon (janvier 2020).
- Membre du jury de la thèse de Jaouad Mourtada, Ecole Polytechnique (juin 2020).
- Président du jury de la thèse de Qiming Du, Sorbonne Université (juin 2020).
- Président du jury de la thèse de Vincent Margot, Sorbonne Université (octobre 2020).
- Président du jury de la thèse d'Alexis Frémond, Université PSL (octobre 2020).
- Président du jury de la thèse de Léa Pannecoucke, Université PSL (octobre 2020).
- Président du jury de la thèse de Johann Faouzi, Sorbonne Université (décembre 2020).
- Président du jury de la thèse de Taieb Touati, Sorbonne Université (février 2021).
- Membre du jury de la thèse d'Ugo Tanielian, Sorbonne Université (avril 2021).
- Président du jury de la thèse d'Aude Sportisse, Sorbonne Université (juin 2021).
- Membre du jury de la thèse de Nicolas Schreuder, Institut Polytechnique de Paris (octobre 2021).
- Membre du jury de la thèse d'Adeline Fermanian, Sorbonne Université (octobre 2021).
- Président du jury de la thèse de Pierre Bertrand, Sorbonne Université (novembre 2021).
- Membre du jury de la thèse de Clément Bénard, Sorbonne Université (décembre 2021).
- Membre du jury de la thèse de Panayotis Papoutsis, Université de Nantes (décembre 2021).
- Membre du jury de la thèse de Jérémie Donà, Sorbonne Université (mai 2022).
- Membre du jury de la thèse d'El Mehdi Saad, Université Paris-Saclay (décembre 2022).
- Membre du jury de la thèse de Paul Freulon, Université de Bordeaux (mars 2023).
- Président du jury de la thèse de Matthieu Kirchmeyer, Sorbonne Université (mai 2023).
- Membre du jury de la thèse d'Ariane Marandon-Carlhian, Sorbonne Université (septembre 2023).
- Membre du jury de la thèse de Ludovic Arnould, Sorbonne Université (octobre 2023).
- Président du jury de la thèse d'Alexandre Godmer, Sorbonne Université (décembre 2023).
- Membre du jury de la thèse de Joachim Bona-Pellissier, Université de Toulouse (décembre 2023).

## Habilitations à diriger des recherches

---

- Rapporteur de l'HDR de Jean-François Dupuy, Université Toulouse III (octobre 2008).
- Rapporteur de l'HDR de Sophie Dabo-Niang, Université Charles-de-Gaulle – Lille 3 (novembre 2009).
- Rapporteur de l'HDR d'Anne-Françoise Yao, Université de la Méditerranée – Aix-Marseille II (novembre 2010).
- Membre du jury de l'HDR d'Agathe Guilloux, Université Paris VI (décembre 2010).
- Président du jury de l'HDR d'Arnaud Guyader, Université Rennes 2 (décembre 2011).
- Membre du jury de l'HDR de Fadoua Balabdaoui, Université Paris-Dauphine (mai 2012).
- Rapporteur de l'HDR de Céline Roget-Vial, Université Claude Bernard Lyon 1 (décembre 2012).
- Rapporteur de l'HDR de Peggy Cénac, Université de Bourgogne (novembre 2013).
- Membre du jury de l'HDR de Thomas Verdebout, Université Charles-de-Gaulle – Lille 3 (novembre 2013).
- Membre du jury de l'HDR d'Olivier Lopez, Université Paris VI (mars 2014).
- Membre du jury de l'HDR de Salim Bouzebda, Université de Technologie de Compiègne (octobre 2014).
- Président du jury de l'HDR de Vianney Perchet, Université Paris Diderot – Paris 7 (décembre 2014).
- Rapporteur de l'HDR de Pierre Pudlo, Université Montpellier 2 (décembre 2014).
- Membre du jury de l'HDR d'Etienne Roquain, Université Paris VI (septembre 2015).
- Membre du jury de l'HDR de Bertrand Michel, Université Paris VI (novembre 2015).
- Membre du jury de l'HDR de Stéphane Gaïffas, Université Paris VI (mars 2016).
- Rapporteur de l'HDR de Yousri Slaoui, Université de Poitiers (octobre 2016).
- Président du jury de l'HDR de Mathieu Emily, Université de Rennes 1 (novembre 2016).
- Président du jury de l'HDR d'Amor Keziou, Université de Reims Champagne-Ardenne (novembre 2016).
- Membre du jury de l'HDR d'Olga Klopp, Université Paris Ouest – Nanterre La Défense (décembre 2016).
- Membre du jury de l'HDR de Patrick Loiseau, EURECOM (décembre 2016).
- Membre du jury de l'HDR d'Alain Celisse, Université de Lille (octobre 2018).
- Membre du jury de l'HDR de Thomas Laloë, Université Nice Sophia Antipolis (novembre 2018).
- Rapporteur de l'HDR de Servane Gey, Université Paris Descartes (février 2019).

- Rapporteur de l'HDR de Mariane Pelletier, Université de Versailles Saint-Quentin-en-Yvelines (mars 2019).
- Rapporteur de l'HDR de Christine Keribin, Université Paris-Sud (novembre 2019).
- Membre du jury de l'HDR de François Portier, Institut Polytechnique de Paris (septembre 2021).
- Président du jury de l'HDR de Thomas Romary, Sorbonne Université (septembre 2021).
- Membre du jury de l'HDR d'Aurélie Fischer, Université Paris Cité (juin 2022).
- Membre du jury de l'HDR de Pavlo Mozharovskyi, Institut Polytechnique de Paris (juillet 2022).
- Membre du jury de l'HDR de Clément Levrard, Université Paris Cité (décembre 2022).
- Président du jury de l'HDR d'Antoine Godichon-Baggioni, Sorbonne Université (mars 2023).
- Membre du jury de l'HDR de Nicolas Obin, Sorbonne Université (septembre 2023).

## Responsabilités diverses

---

- Rapporteur d'articles pour les revues *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, *Annales de l'ISUP*, *Annals of Statistics*, *Applied and Computational Harmonic Analysis*, *Climate Research*, *Comptes Rendus de l'Académie des Sciences*, *Computational Statistics*, *Computational Statistics & Data Analysis*, *Electronic Journal of Statistics*, *ESAIM: Probability and Statistics*, *IEEE Transactions on Information Theory*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *IEEE Transactions on Signal Processing*, *Institute of Mathematical Statistics*, *Journal de la Société Française de Statistique*, *Journal of Agricultural, Biological, and Environmental Statistics*, *Journal of Machine Learning Research*, *Journal of Multivariate Analysis*, *Journal of Nonparametric Statistics*, *Journal of Statistical Computation and Simulation*, *Journal of Statistical Planning and Inference*, *Journal of the American Statistical Association*, *Journal of the Royal Statistical Society: Series B*, *Machine Learning*, *Mathematical Methods of Statistics*, *PNAS*, *Statistics & Decisions*, *Statistical Inference for Stochastic Processes*, *TEST*, *The Canadian Journal of Statistics*, *Transactions on Knowledge Discovery from Data*, *The Canadian Journal of Statistics*, *Transactions on Modeling and Computer Simulation*.
- Editeur régional pour la France et la Suisse de la revue *Statistical Theory & Methods Abstracts* (2002-2004).
- Membre titulaire nommé des Commissions de spécialistes de l'Université Paul-Valéry – Montpellier III (sections 25-26-27, 2004-2008), de l'Université Toulouse 1 (section 26, 2006-2008) et de l'INSA Toulouse (sections 25-26, 2006-2008).
- Responsable du M1 commun aux deux spécialités mathématiques et statistique du master Informatique-Mathématiques-Statistique de l'Université Montpellier 2 (2004-2007).
- Responsable du Séminaire de Probabilités et Statistique de l'Université Montpellier 2 (2005-2007).
- Rapporteur de projets pour l'Agence Nationale de la Recherche (ANR), Programme *Jeunes chercheuses et jeunes chercheurs* et Programme *Blanc*.

- Editeur associé du *Journal de la Société Française de Statistique* (2007-2008).
- Membre du jury du *prix du Docteur Norbert Marx* décerné par la Société Française de Statistique (2007).
- Coresponsable avec Michel Broniatowski (Professeur, Sorbonne Université), Michel Delecroix (Professeur honoraire, Sorbonne Université) puis Olivier Lopez du Séminaire du LSTA (2007-2010).
- Coresponsable avec Stéphane Boucheron (Professeur, Université Paris Cité), Alexandre Tsybakov (Professeur, ENSAE) puis Pierre Alquier du Groupe de travail de Statistique (2007-2010).
- Coresponsable du Séminaire SMILE (Statistical MachIne LEarning) in Paris (2007-2018).
- Membre élu au Conseil de l'Institut de Statistique de l'Université de Paris (2007-2013).
- Directeur des études de l'Institut de Statistique de l'Université de Paris (2007-2012).
- Membre de missions d'expertise de l'Agence d'Evaluation de la Recherche et de l'Enseignement Supérieur (AERES, une mission en 2008, deux missions en 2009, une mission en 2010 et une mission en 2012).
- Missions d'évaluation à la demande de la City University de Hong Kong, du Ministère de l'Agriculture et de la Pêche français, du CNRS, de l'INRIA et du COFECUB (Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil).
- Membre du jury du *prix de thèse SMAI-MAS en l'honneur de Jacques Neveu* décerné par la Société de Mathématiques Appliquées et Industrielles (2009-2011).
- Membre des Comités de sélection de l'Université Paris VI (depuis 2008), de l'Université Paris-Dauphine (section 26, 2009-2011 et 2014), de l'Université Toulouse 1 (section 26, 2008-2010), de l'INSA Rennes (section 26, 2010), de l'Université Paris Diderot – Paris 7 (section 26, 2011), du CNAM (section 26, 2011, 2012 et 2015), de l'Université Claude Bernard Lyon 1 (section 26, 2012), de l'Université Joseph Fourier (section 26, 2013), de l'Université Paris-Est Marne-la-Vallée (section 26, 2018), de l'Université Paris-Sud (section 26, 2019), de l'Université Grenoble Alpes (section 26, 2020), de l'Université de Lille (section 26, 2021), de l'Université Rennes 2 (section 27, 2022) et de l'Université Toulouse III - Paul Sabatier (section 26, 2023).
- Membre de jurys de recrutement à l'Université Libre de Bruxelles (2009 et 2014), à AgroParisTech (2009 et 2013), à l'Ecole des Ponts ParisTech (2018), à l'Ecole Polytechnique (2019-2022, 2024), à l'ENSAI (2019), à l'Université de Zurich (2022) et à l'Université Paris-Panthéon-Assas (2022).
- Membre élu au Conseil de l'UFR de Mathématiques de l'Université Paris VI (2009-2013).
- Membre nommé au Conseil de l'Ecole Doctorale Pierre Louis de Santé Publique à Paris (depuis 2009).
- Editeur associé de l'*International Statistical Review* (depuis 2009).
- Editeur associé des *Annales de l'ISUP* (2010-2022).
- Membre nommé au Conseil de la Recherche du CREST (Centre de Recherche en Economie et Statistique, 2010-2012) puis au Conseil Scientifique du GENES (Groupe des Ecoles Nationales d'Economie et Statistique, 2012-2017).

- Membre du jury du *prix Marie-Jeanne Laurent-Duhamel* décerné par la Société Française de Statistique (2011, 2012, 2014 et 2015).
- Editeur associé (2011-2013) puis coéditeur en chef (2013-2017) de la revue *ESAIM: Probability and Statistics*.
- Directeur adjoint du master Mathématiques et Applications de l'Université Paris VI (2011-2013).
- Membre élu au Conseil National des Universités (CNU), section 26 (2011-2015).
- Membre nommé au Conseil Scientifique de la Société Mathématique de France (2012-2015).
- Membre nommé à la commission d'attribution de la prime d'encadrement doctoral et de recherche de l'Université Paris VI (2012) puis de Sorbonne Université (2018).
- Membre du jury du *prix Pierre-Simon de Laplace* décerné par la Société Française de Statistique (2013).
- Directeur du Laboratoire de Statistique Théorique et Appliquée de l'Université Paris VI (2013-2018).
- Membre nommé au Conseil Scientifique de l'UFR de Mathématiques de l'Université Paris VI (depuis 2013).
- Membre du comité éditorial de la collection *Mathematics and Statistics* de l'éditeur ISTE-Hermes-Wiley (depuis 2013).
- Président du comité d'évaluation d'une mission d'expertise de l'AERES (2013) puis du Hcéres (2019, deux missions).
- Examineur au second concours d'entrée à l'Ecole Normale Supérieure de Rennes, épreuve d'entretien avec le jury (2014).
- Formateur auprès des professeurs de classes préparatoires aux grandes écoles dans le cadre des *Journées de Probabilités de l'ENS* (2014).
- Editeur associé du *Journal of the American Statistical Association* (depuis 2017).
- Editeur associé de la revue *Biometrika* (depuis 2018).
- Directeur adjoint du Laboratoire de Probabilités, Statistique et Modélisation de Sorbonne Université (2018-2021).
- Membre du comité éditorial de la collection *Mathématiques et Applications* de l'éditeur SMAI-Springer (depuis 2018).
- Conseiller de la Vice-Présidente Recherche, Innovation et Science Ouverte, en charge de l'intelligence artificielle (2018-2020).
- Editeur associé de la revue *The Annals of Statistics* (depuis 2019).
- Représentant de Sorbonne Université au sein du Flagship 3 (*Numérisation, modélisation, transformations*) de l'alliance universitaire européenne 4EU+ (depuis 2019).
- Directeur du Sorbonne Center for Artificial Intelligence (SCAI, depuis 2019).



- Coresponsable avec Patrick Gallinari et Maxime Sangnier du parcours *Apprentissage et Algorithmes* (M2A) du master Mathématiques et Applications de Sorbonne Université (depuis 2019).
- Membre nommé au comité d'experts chargé de la Synthèse Nationale et de Prospective sur les Mathématiques, mis en place par le Hcéres (2020-2022).
- Membre extérieur nommé au Conseil Scientifique et de Prospective de l'Institut de Mathématiques de Toulouse (depuis 2020).
- Membre nommé au Comité Scientifique et Technologique de Programme du PEPR sur l'IA (depuis 2022).

## Relations avec le monde industriel

---

- Membre élu au Conseil de la Société Française de Statistique (2005-2009, 2010-2014 puis 2015-2018). Elu vice-président de l'association en juin 2006, réélu en juin 2008, juin 2010, juin 2011 et juin 2013. Elu président de juin 2015 à juin 2018.
- Membre du Comité Scientifique du GdR MASCOT-NUM (Groupement de Recherche Méthodes d'Analyse Stochastique pour les COdes et Traitements NUMériques, 2008-2020).
- Conseil scientifique auprès de plusieurs cabinets d'avocats (depuis 2012).
- Membre élu de l'International Statistical Institute (2015).
- Dépôt d'un brevet intitulé *Procédé d'analyse d'un vol d'un aéronef pour évaluer l'exposition à un risque donné dudit vol*, en partenariat avec la société Safety Line (n° 15/51975, 2015).
- Co-porteur avec Raphaël Porcher (Professeur-Praticien hospitalier, Université Paris Cité) du projet ANR OPT (*Optimization of Personalized Treatments*, Programme AAPG 2018, 2019-2021, budget : 290 304 euros).
- Représentant du Président de Sorbonne Université au Conseil de la Fondation La Science Statistique (depuis 2020).
- Membre nommé à l'*IMS Committee on Nominations* (Institute of Mathematical Statistics, 2020-2021).
- Directeur adjoint de l'Unité ELLIS Paris (*European Laboratory for Learning and Intelligent Systems*, depuis 2020).
- Membre élu au Conseil d'Administration de Cap Digital (représentant de Sorbonne Université, depuis 2021).

## Récompenses scientifiques

---

- *Förderpreis* décerné par le Helmholtz-Zentrum Geesthacht en 1997 : ce prix était destiné à récompenser un travail de recherche original effectué au sein du Helmholtz-Zentrum Geesthacht durant l'année 1996-1997.
- *Prix Marie-Jeanne Laurent-Duhamel* décerné par la Société Française de Statistique en 2003 : ce prix honore tous les deux ans les travaux de recherche d'un jeune statisticien francophone.

- Nomination à l’Institut Universitaire de France en qualité de membre junior (promotion 2012).
- Prix du meilleur article publié en 2014-2015 par les Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques pour Biau, G., Cérou, F. et Guyader, A. (2015). *New insights into Approximate Bayesian Computation*, **Vol. 51**, pp. 376–403.
- Prix Michel Monpetit – Inria décerné par l’Académie des sciences en 2018 : ce prix récompense tous les ans un chercheur ou un ingénieur pour ses travaux dans le domaine des mathématiques appliquées ou de l’informatique.

## Publications

---

### Livres

- Biau, G., Droniou, J. et Herzlich, M. (2010). Mathématiques et statistique pour les sciences de la nature — Modéliser, comprendre et appliquer, *EDP Sciences*, 531 pages, ISBN 978-2-7598-0481-8.
- Biau, G. et Devroye, L. (2015). Lectures on the nearest neighbor method, *Springer*, Cham, 290 pages, ISBN 978-3-319-25386-2.

### Articles parus ou à paraître

- [1] Biau, G. (1999). Estimateurs à noyau itérés : synthèse bibliographique, *Journal de la Société Française de Statistique*, **Vol. 1**, pp. 41–67.
- [2] Biau, G., Zorita, E., von Storch, H. et Wackernagel, H. (1999). Estimation of precipitation by kriging in the EOF space of the Sea Level Pressure field, *Journal of Climate*, **Vol. 12**, pp. 1070–1085.
- [3] Berlinet, A. et Biau, G. (2001). Iterated Barron density estimators, *Comptes Rendus de l’Académie des Sciences*, **Vol. 332**, pp. 459–464.
- [4] Berlinet, A. et Biau, G. (2001). Minimax bounds in nonparametric estimation of multidimensional deterministic dynamical systems, *Statistical Inference for Stochastic Processes*, **Vol. 4**, pp. 229–248.
- [5] Abraham, C., Biau, G. et Cadre, B. (2002). Chaotic properties of mappings on a probability space, *Journal of Mathematical Analysis and Applications*, **Vol. 266**, pp. 420–431.
- [6] Beirlant, J., Berlinet, A., Biau, G. et Vajda, I. (2002). Divergence-type errors of smooth Barron-type density estimators, *TEST*, **Vol. 11**, pp. 191–217.
- [7] Biau, G. (2002). Optimal asymptotic quadratic errors of density estimators on random fields, *Statistics & Probability Letters*, **Vol. 60**, pp. 297–307.
- [8] Abraham, C., Biau, G. et Cadre, B. (2003). Simple estimation of the mode of a multivariate density, *The Canadian Journal of Statistics*, **Vol. 31**, pp. 23–34.
- [9] Biau, G. et Devroye, L. (2003). On the risk of estimates for block decreasing densities, *Journal of Multivariate Analysis*, **Vol. 86**, pp. 143–165.
- [10] Biau, G. (2003). Spatial kernel density estimation, *Mathematical Methods of Statistics*, **Vol. 12**, pp. 371–390.

- [11] Abraham, C., Biau, G. et Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode, *ESAIM: Probability and Statistics*, **Vol. 8**, pp. 1–11.
- [12] Abraham, C., Biau, G. et Cadre, B. (2004). On Lyapunov exponent and sensitivity, *Journal of Mathematical Analysis and Applications*, **Vol. 290**, pp. 395–404.
- [13] Biau, G. et Devroye, L. (2004). A note on density model size testing, *IEEE Transactions on Information Theory*, **Vol. 50**, pp. 576–581.
- [14] Biau, G. (2004). Estimation de la densité et tests par la méthode combinatoire pénalisée, *Journal de la Société Française de Statistique*, **Vol. 4**, pp. 5–24.
- [15] Berline, A. et Biau, G. (2004). Iterated modified histograms as dynamical systems, *Journal of Nonparametric Statistics*, **Vol. 16**, pp. 385–401.
- [16] Biau, G. et Cadre, B. (2004). Nonparametric spatial prediction, *Statistical Inference for Stochastic Processes*, **Vol. 7**, pp. 327–349.
- [17] Biau, G. et Devroye, L. (2005). Density estimation by the penalized combinatorial method, *Journal of Multivariate Analysis*, **Vol. 94**, pp. 196–208.
- [18] Berline, A., Biau, G. et Rouvière, L. (2005). Parameter selection in modified histogram estimates, *Statistics*, **Vol. 39**, pp. 91–105.
- [19] Biau, G. et Wegkamp, M. (2005). A note on minimum distance estimation of copula densities, *Statistics & Probability Letters*, **Vol. 73**, pp. 105–114.
- [20] Biau, G., Bunea, F. et Wegkamp, M.H. (2005). Functional classification in Hilbert spaces, *IEEE Transactions on Information Theory*, **Vol. 51**, pp. 2163–2172.
- [21] Berline, A., Biau, G. et Rouvière, L. (2005). Optimal  $L_1$  bandwidth selection for variable kernel density estimates, *Statistics & Probability Letters*, **Vol. 74**, pp. 116–128.
- [22] Biau, G. et Györfi, L. (2005). On the asymptotic properties of a nonparametric  $L_1$ -test statistic of homogeneity, *IEEE Transactions on Information Theory*, **Vol. 51**, pp. 3965–3973.
- [23] Abraham, C., Biau, G. et Cadre, B. (2006). On the kernel rule for function classification, *Annals of the Institute of Statistical Mathematics*, **Vol. 58**, pp. 619–633.
- [24] Biau, G. et Bleakley, K. (2006). Statistical inference on graphs, *Statistics & Decisions*, **Vol. 24**, pp. 209–232.
- [25] Bleakley, K., Giudicelli, V., Wu, J., Lefranc, M.-P. et Biau, G. (2006). IMG-T standardization for statistical analyses of T cell receptor junctions: The TRAV-TRAJ example, *In Silico Biology*, **Vol. 6**, pp. 573–588.
- [26] Biau, G., Cadre, B. et Pelletier, B. (2007). A graph-based estimator of the number of clusters, *ESAIM: Probability and Statistics*, **Vol. 11**, pp. 272–280.
- [27] Bleakley, K., Biau, G. et Vert, J.-P. (2007). Supervised reconstruction of biological networks with local models, *Bioinformatics*, **Vol. 23**, pp. i57–i65.
- [28] Biau, G., Biau, O. et Rouvière, L. (2007). Nonparametric forecasting of the manufacturing output growth with firm-level survey data, *Journal of Business Cycle Measurement and Analysis*, **Vol. 3**, pp. 317–332.

- [29] Biau, G., Devroye, L. et Lugosi, G. (2008). On the performance of clustering in Hilbert spaces, *IEEE Transactions on Information Theory*, **Vol. 54**, pp. 781–790.
- [30] Beirlant, J., Berlinet, A. et Biau, G. (2008). Higher order estimation at Lebesgue points, *Annals of the Institute of Statistical Mathematics*, **Vol. 60**, pp. 651–677.
- [31] Biau, G., Cadre, B., Devroye, L. et Györfi, L. (2008). Strongly consistent model selection for densities, *TEST*, **Vol. 17**, pp. 531–545.
- [32] Biau, G., Cadre, B. et Pelletier, B. (2008). Exact rates in density support estimation, *Journal of Multivariate Analysis*, **Vol. 99**, pp. 2185–2207.
- [33] Berlinet, A., Biau, G. et Rouvière, L. (2008). Functional supervised classification with wavelets, *Annales de l'ISUP*, **Vol. 52**, pp. 61–80.
- [34] Biau, G., Devroye, L. et Lugosi, G. (2008). Consistency of random forests and other averaging classifiers, *Journal of Machine Learning Research*, **Vol. 9**, pp. 2015–2033.
- [35] Bleakley, K., Lefranc, M.-P. et Biau, G. (2008). Recovering probabilities for nucleotide trimming processes for T cell receptor TRA and TRG V-J junctions analysed with IMGT tools, *BMC Bioinformatics*, **Vol. 9**, pp. 408–414.
- [36] Biau, G., Cadre, B., Mason, D.M. et Pelletier, B. (2009). Asymptotic normality in density support estimation, *Electronic Journal of Probability*, **Vol. 14**, pp. 2617–2635.
- [37] Biau, G., Cérou, F. et Guyader, A. (2010). On the rate of convergence of the bagged nearest neighbor estimate, *Journal of Machine Learning Research*, **Vol. 11**, pp. 687–712.
- [38] Biau, G., Bleakley, K., Györfi, L. et Ottucsák, G. (2010). Nonparametric sequential prediction of time series, *Journal of Nonparametric Statistics*, **Vol. 22**, pp. 297–317.
- [39] Biau, G., Cadre, B. et Rouvière, L. (2010). Statistical analysis of  $k$ -nearest neighbor collaborative recommendation, *The Annals of Statistics*, **Vol. 38**, pp. 1568–1592.
- [40] Biau, G., Cérou, F. et Guyader, A. (2010). Rates of convergence of the functional  $k$ -nearest neighbor estimate, *IEEE Transactions on Information Theory*, **Vol. 56**, pp. 2034–2040.
- [41] Biau, G. et Devroye, L. (2010). On the layered nearest neighbour estimate, the bagged nearest neighbour estimate and the random forest method in regression and classification, *Journal of Multivariate Analysis*, **Vol. 101**, pp. 2499–2518.
- [42] Biau, G. et Patra, B. (2011). Sequential quantile prediction of time series, *IEEE Transactions on Information Theory*, **Vol. 57**, pp. 1664–1674.
- [43] Biau, G., Chazal, F., Cohen-Steiner, D., Devroye, L. et Rodríguez, C. (2011). A weighted  $k$ -nearest neighbor density estimate for geometric inference, *Electronic Journal of Statistics*, **Vol. 5**, pp. 204–237.
- [44] Biau, G. et Fischer, A. (2012). Parameter selection for principal curves, *IEEE Transactions on Information Theory*, **Vol. 58**, pp. 1924–1939.
- [45] Biau, G. et Mas, A. (2012). PCA-kernel estimation, *Statistics & Risk Modeling*, **Vol. 29**, pp. 19–46.
- [46] Biau, G. (2012). Analysis of a random forests model, *Journal of Machine Learning Research*, **Vol. 13**, pp. 1063–1095.

- [47] Biau, G. et Yatracos, Y.G. (2012). On the shrinkage estimation of variance and Pitman closeness criterion, *Journal de la Société Française de Statistique*, **Vol. 153**, pp. 5–21.
- [48] Biau, G., Devroye, L., Dujmović, V. et Krzyżak, A. (2012). An affine invariant  $k$ -nearest neighbor regression estimate, *Journal of Multivariate Analysis*, **Vol. 112**, pp. 24–34.
- [49] Alquier, P. et Biau, G. (2013). Sparse single-index model, *Journal of Machine Learning Research*, **Vol. 14**, pp. 243–280.
- [50] Biau, G. et Devroye, L. (2013). Cellular tree classifiers, *Electronic Journal of Statistics*, **Vol. 7**, pp. 1875–1912.
- [51] Kruppa, J., Liu, Y., Biau, G., Kohler, M., König, I.R., Malley, J.D. et Ziegler, A. (2014). Probability estimation with machine learning methods for dichotomous and multicategory outcome: Theory, *Biometrical Journal*, **Vol. 56**, pp. 534–563.
- [52] Biau, G., Cérou, F. et Guyader, A. (2015). New insights into Approximate Bayesian Computation, *Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques*, **Vol. 51**, pp. 376–403.
- [53] Biau, G. et Mason, D.M. (2015). High-dimensional  $p$ -norms, in *Mathematical Statistics and Limit Theorems: Festschrift in Honour of Paul Deheuvels*, ed. Hallin, M., Mason, D.M., Pfeifer, D. et Steinebach, J.G., pp. 21–40, Springer, Cham.
- [54] Scornet, E., Biau, G. et Vert, J.-P. (2015). Consistency of random forests, *The Annals of Statistics*, **Vol. 43**, pp. 1716–1741.
- [55] Biau, G., Cadre, B. et Paris, Q. (2015). Cox process functional learning, *Statistical Inference for Stochastic Processes*, **Vol. 18**, pp. 257–277.
- [56] Biau, G., Fischer, A., Guedj, B. et Malley, J.D. (2016). COBRA: A combined regression strategy, *Journal of Multivariate Analysis*, **Vol. 146**, pp. 18–28.
- [57] Biau, G. et Scornet, E. (2016). A random forest guided tour (with comments and a rejoinder by the authors), *TEST*, **Vol. 25**, pp. 197–227.
- [58] Biau, G., Bleakley, K. et Mason, D.M. (2016). Long signal change-point detection, *Electronic Journal of Statistics*, **Vol. 10**, pp. 2097–2123.
- [59] Biau, G., Bleakley, K. et Cadre, B. (2016). The statistical performance of collaborative inference, *Journal of Machine Learning Research*, **Vol. 17(62)**, pp. 1–29.
- [60] Biau, G. et Zenine, R. (2018). Online asynchronous distributed regression, *Annales de l’ISUP*, **Vol. 62**, pp. 29–58.
- [61] Biau, G., Cadre, B. et Rouvière, L. (2019). Accelerated gradient boosting, *Machine Learning*, **Vol. 108**, pp. 971–992.
- [62] Biau, G., Scornet, E. et Welbl, J. (2019). Neural random forests, *Sankhyā A*, **Vol. 81**, pp. 347–386.
- [63] Biau, G., Cadre, B., Sangnier, M. et Tanielian, U. (2020). Some theoretical properties of GANs, *The Annals of Statistics*, **Vol. 48**, pp. 1539–1566.
- [64] Bénard, C., Biau, G., Da Veiga, S. et Scornet, E. (2021). SIRUS: Stable and Interpretable RRule Set for classification, *Electronic Journal of Statistics*, **Vol. 15**, pp. 427–505.

- [65] Tanielian, U., Sangnier, M. et Biau, G. (2021). Approximating Lipschitz continuous functions with GroupSort neural networks, in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ed. Banerjee, A. et Fukumizu, K., Proceedings of Machine Learning Research, **Vol. 130**, pp. 442–450, PMLR.
- [66] Bénard, C., Biau, G., Da Veiga, S. et Scornet, E. (2021). Interpretable random forests via rule extraction, in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ed. Banerjee, A. et Fukumizu, K., Proceedings of Machine Learning Research, **Vol. 130**, pp. 937–945, PMLR.
- [67] Du, Q., Biau, G., Petit, F. et Porcher, R. (2021). Wasserstein random forests and applications in heterogeneous treatment effects, in *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, ed. Banerjee, A. et Fukumizu, K., Proceedings of Machine Learning Research, **Vol. 130**, pp. 1729–1737, PMLR.
- [68] Biau, G., Sangnier, M. et Tanielian, U. (2021). Some theoretical insights into Wasserstein GANs, *Journal of Machine Learning Research*, **Vol. 22**(119), pp. 1–45.
- [69] Biau, G. et Cadre, B. (2021). Optimization by gradient boosting, in *Advances in Contemporary Statistics and Econometrics: Festschrift in Honor of Christine Thomas-Agnan*, ed. Daouia, A. et Ruiz-Gazen, A., pp. 23–44, Springer, Cham.
- [70] Fermanian, A., Marion, P., Vert, J.-P. et Biau, G. (2021). Framing RNN as a kernel method: A neural ODE approach, in *Advances in Neural Information Processing Systems*, ed. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P.S. et Wortman Vaughan, J., **Vol. 34**, pp. 3121–3134, Curran Associates, Inc. (oral presentation).
- [71] Bénard, C., Biau, G., Da Veiga, S. et Scornet, E. (2022). SHAFF: Fast and consistent SHAPley eFFect estimates via random Forests, in *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, ed. Camps-Valls, G., Ruiz, F.J.R. et Valera, I., Proceedings of Machine Learning Research, **Vol. 151**, pp. 5563–5582, PMLR.
- [72] Fermanian, A., Chang, J., Lyons, T. et Biau, G. (2024). The insertion method to invert the signature of a path, *Festschrift in Honor of Marc Hallin*, à paraître.
- [73] Stéphanovitch, A., Tanielian, U., Cadre, B., Klutchnikoff, N. et Biau, G. (2024). Optimal 1-Wasserstein distance for WGANs, *Bernoulli*, à paraître.
- [74] Marion, P., Wu, Y.-H., Sander, M.E. et Biau, G. (2024). Implicit regularization of deep residual networks towards neural ODEs, *International Conference on Learning Representations*, à paraître (spotlight presentation).

## Articles soumis

- [75] Marion, P., Fermanian, A., Biau, G. et Vert, J.-P. (2022). Scaling ResNets in the large-depth regime, soumis à la revue *Journal of Machine Learning Research*.
- [76] Doumèche, N., Biau, G. et Boyer, C. (2023). On the convergence of PINNs, soumis à la revue *Bernoulli*.
- [77] Doumèche, N., Bach, F., Biau, G. et Boyer, C. (2024). Physics-informed machine learning as a kernel method, soumis à la conférence *COLT 2024*.

## Actes de conférences, chapitres de livres et rapports de recherche

- [a] Biau, G. (1999). Downscaling of precipitation combining kriging and Empirical Orthogonal Function analysis, in *GeoENV II – Geostatistics for Environmental Applications*, ed. Gómez-Hernández, J., Soares, A. et Froidevaux, R., pp. 151–162, Kluwer, Dordrecht.
- [b] Berline, A. et Biau, G. (1999). A chaotic non-uniform random variate generator, *Université Montpellier 2, Rapport de Recherche 99-06*.
- [c] Berline, A. et Biau, G. (2002). Estimation de densité et prise de décision, in *Décision et Reconnaissance de Formes en Signal*, ed. Lengellé, R., pp. 141–179, Hermès, Paris.
- [d] Biau, G. et Györfi, L. (2006). On a  $L_1$ -test statistic of homogeneity, *Proceedings of the BMS Conference BeNeLuxFra2005, Bulletin of the Belgian Mathematical Society Simon Stevin*, **Vol. 13**, pp. 877–881.
- [e] Biau, G. et Devroye, L. (2014). Cellular tree classifiers, in *Algorithmic Learning Theory: Proceedings of the 25th International Conference, ALT 2014*, ed. Auer, P., Clark, A., Zeugmann, T. et Zilles, S., Lecture Notes in Computer Science, **Vol. 8776**, pp. 8–17, Springer, Cham.
- [f] Biau, G. et Levrard, C. (2016). Comments on: *Probability enhanced effective dimension reduction for classifying sparse functional data* by Yao, F., Wu, Y. et Zou, J., *TEST*, **Vol. 25**, pp. 41–43.
- [g] Biau, G. et Massart, P. (2017). Science des données : naissance ou renaissance ?, in *Les Big Data à Découvert*, ed. Bouzeghoub, M. et Mosseri, R., pp. 134–135, CNRS Editions, Paris.
- [h] Biau, G. et Fermanian, A. (2020). Learning with signatures, in *Functional and High-Dimensional Statistics and Related Fields*, ed. Aneiros, G., Horová, I., Hušková, M. et Vieu, P., pp. 19–26, Springer, Cham.