

## Synthèse de carrière

J'ai fait mes études supérieures à l'**Ecole Normale Supérieure de Cachan**, où j'ai suivi **magistère mathématiques-informatique** (co-habilité avec l'Université Paris 7 Diderot) avant de passer l'**agrégation de mathématiques**. J'ai suivi le DEA de modélisation stochastique et statistique de l'Université Paris XI Orsay, où j'ai ensuite effectué ma **thèse sous la direction d'Elisabeth Gassiat et de Stéphane Boucheron**. Mes travaux portaient sur la théorie de l'information, plus précisément sur le codage universel et le principe du Minimum Description Length (MDL) : il s'agissait d'utiliser des idées récentes de statistique mathématique (comme en sélection de modèles) pour les appliquer dans un contexte de codage (par exemple, pour le codage sur les alphabets très grands voire infinis), et réciproquement d'utiliser les techniques de théorie de l'information pour montrer des propriétés statistiques (par exemple, la convergence d'estimateurs BIC dans les modèles markoviens à longueur variable).

J'ai ensuite effectué une année en tant qu'**ATER à l'ENS Cachan**, en grande partie partagée entre l'achèvement de mes travaux de thèse et la préparation des étudiants à l'agrégation. A l'issue de cette année, j'ai été recruté en tant que **chargé de recherche CNRS** sur un poste croisé (section 1 -> 7) et affecté au LTCI de Telecom ParisTech, dans l'équipe STA d'Eric Moulines et Olivier Cappé. Dans cet environnement très dynamique, mes thématiques de recherche ont progressivement évolué vers l'apprentissage statistique, tout en gardant une composante importante autour des modèles markoviens. A partir de 2008, j'ai co-encadré la thèse de Sarah Filippi (thèse Orange, avec un volet applicatif nouveau pour moi), puis à celle d'Emilie Kaufmann (thèse purement académique récompensée par le Prix Jacques Neveu 2014). Mes cinq années passées à Telecom m'ont également permis de bien connaître le système des grandes écoles, notamment à travers des enseignements que j'ai effectués dans plusieurs établissements (Telecom bien sûr mais aussi Centrale, Ecole des Ponts, ENS Cachan, HEC Paris). J'ai participé pendant deux ans aux activités naissantes de l'Institut des Sciences de la Communication du CNRS (ISCC), une expérience mitigée mais intéressante. J'ai par ailleurs participé au jury d'entrée de l'Ecole Normale Supérieure voie BL entre 2009 et 2012, en tant que responsable des épreuves écrites et orales de mathématiques (de cette expérience est issue un livre, voir ci-dessous).

En novembre 2011, j'ai reçu mon Habilitation à Diriger des Recherches. J'ai été recruté l'année suivante comme **professeur à l'Université Paul Sabatier, dans l'Institut de Mathématiques de Toulouse**, où j'exerce depuis. J'y poursuis mes travaux, en m'attachant notamment à développer l'interface entre mathématiques et informatique. J'y ai encadré trois thèses déjà soutenues, et d'autres sont en cours. Je porte un projet ANR (en collaboration avec Antoine Chambaz de Paris 10) et je suis responsable à l'UPS d'un autre projet ANR (porté par Bogdan Cautis de Paris Sud). Sur le plan des enseignements, j'ai pris dès mon arrivée la responsabilité de la Licence et du Master Statistique et Informatique Décisionnelle (SID). J'ai monté au printemps suivant le dossier pour transformer la formation en Coursus de Master en Ingénierie (CMI). Le réseau Figure a validé le projet à partir de septembre 2013, et m'a demandé de parti-

ciper aux expertises des autres CMI à partir de 2015. Depuis septembre 2017, **je dirige le Département de Mathématiques**. Sur le plan de la recherche, je suis **Directeur Adjoint du GDR MaDICS** (Masses de Données, Informations et Connaissances en Sciences) du CNRS après avoir participé à son montage. Son objectif est d'animer la recherche pluridisciplinaire sur les masses de données, notamment en suscitant des collaborations entre spécialistes ayant insuffisamment l'occasion ou l'habitude de travailler ensemble. Sur un plan plus local, j'ai participé au comité exécutif du Labex CIMI (dont j'ai démissionné en prenant la direction du département). Via l'animation d'un groupe de travail consacré à l'apprentissage machine puis via l'organisation du Trimestre Thématique sur ce thème organisé entre septembre et décembre 2015, j'essaie avec mes collègues intéressés de l'IMT et de l'IRIT d'**animer la recherche en Machine Learning** ; j'ai créé avec eux l'**équipe-projet AOC** (Apprentissage, Optimisation, Complexité) regroupant plusieurs chercheurs d'équipes différentes de ces deux laboratoires, dont j'ai la responsabilité.

## Activités pédagogiques

### 1. Présentation de l'activité d'enseignement

Mes enseignements à l'Université Paul Sabatier portent majoritairement sur l'**aléatoire**, en probabilités, statistique mathématique, statistique appliquée et apprentissage statistique. Cela répond en effet au besoin du laboratoire, qui compte relativement peu de collègues à même d'intervenir en particulier sur les trois derniers de ces quatre disciplines. Ils se composent bien sûr à la fois de cours magistraux, de TD et des TP. Je suis toutefois très attaché à enseigner, dès que possible, d'autres parties des mathématiques (j'ai notamment pris la responsabilité d'un cours d'introduction à l'arithmétique et à l'algorithmique en Licence 1).

Mon engagement dans les enseignements a d'abord été très fortement marqué par la reprise de la **responsabilité de la licence et du master SID** (Statistique et Informatique Décisionnelle) dès novembre 2012, qui n'avait pas été anticipée. Je me suis dès lors beaucoup mobilisé pour en maîtriser les contenus, mais aussi la culture et les équilibres particuliers. Cette formation, ancien IUP en grand besoin de renouvellement, présente en effet la particularité d'être exactement à cheval entre les départements d'informatique et de mathématique, aux habitudes souvent divergentes. Il avait en outre un recrutement extrêmement fragile, presque entièrement externe à l'UPS en Licence 3 et donc très volatile (les étudiants provenant alors principalement de tous les IUT STID de France : Paris, Lisieux, Avignon, etc.). Cette volatilité était d'autant plus grande que, s'il avait longtemps joui d'un positionnement singulier lui donnant un quasi monopole sur son créneau, il est depuis quelques années de plus en plus concurrencé par nombre de nouvelles formations nées dans la vague du *big data* un peu partout en France.

Il y avait donc beaucoup à faire, et comme j'arrivais d'une école d'ingénieur je tenais à ce que la qualité de formation offerte aux étudiants de la formation soit au moins au même niveau. Dès la première année, à la demande du département de mathématique et avec l'aide du département d'informatique, j'ai **monté le dossier** auprès du réseau FIGURE afin que la formation devienne un Coursus de Master en Ingénierie (**CMI**). Cela nécessitait de réviser des contenus (dans une thématique en forte évolution), d'ajouter deux années initiales avec des créations de cours (le CMI étant un cursus de 5 ans aux équilibres définis par un cahier des charges contraignant), de développer

la formation par la recherche, etc. Mon objectif a toujours été de conserver les forces de la formation (son excellente insertion professionnelle, due à son positionnement stratégique, et son très bon réseau d'entreprises partenaires alimenté en partie par des anciens élèves) tout en adaptant les pratiques et les contenus aux exigences de ses nouvelles ambitions. Dans cet objectif, j'ai mis en place un **partenariat** avec l'école de commerce Toulouse Business School (**TBS**) pour l'accueil d'un nombre limité de leurs étudiants les plus motivés dans la licence 3 (en échange notamment de la possibilité, pour quelques uns de nos étudiants, d'être accueillis gratuitement pour un semestre de M2 à TBS). En outre, j'ai avec ma collègue Cécile Chouquet introduit pour la rentrée 2016 l'**alternance en Master 2**, sous la forme d'un contrat de professionnalisation.

Le montage et l'animation du CMI m'ont apporté une grande expérience non seulement avec les étudiants et les équipes pédagogiques, mais aussi de l'administration de l'université et même du système universitaire au niveau national, grâce au réseau Figure. Ce dernier organise en effet des rencontres permettant aux porteurs de CMI de se rencontrer et de travailler ensemble sur des ateliers. Mais il m'a aussi permis de passer de l'autre côté de la barrière, en allant **expertiser** deux projets de CMI mathématique-informatique, et en faisant la visite à deux ans d'un autre CMI — expériences précieuses pour avoir conscience des similarités et des différences qui existent entre les établissements français. Concernant les éléments pédagogiques, une des innovations les plus visibles que j'ai introduites dans la formation est le nouveau **projet "chef-d'œuvre"** sur lequel travaillent à plein temps pendant deux semaines (et plus longtemps en amont) les trois promotions de L3, M1 et M2. Ce projet, porté chaque année par (au minimum) un duo d'enseignants-chercheurs mathématicien-informaticien, comporte non seulement des aspects techniques de synthèse sur la gestion et l'analyse des données, mais également de véritables défis d'organisation collective et de gestion de projet qui sont confiés aux étudiants eux-mêmes (des responsables sont désignés, des groupes sont formés avec des rôles, etc.). La première année, j'ai préparé et encadré le projet avec Karen Pinel-Sauvagnat (IRIT) : il portait sur les données ouvertes de l'opérateur de transport en commun Tisseo. Cette année, c'est Riad Mokadem, Dominique Bontemps et Jonathan Louëdec qui ont encadré un projet de prototype avec pour "client" une société de taxi. Ce projet a valu à la formation une publicité dans les médias qui a dépassé nos attentes, et des discussions sont en cours avec la vice-présidence de l'université à la valorisation, Toulouse-Tech Transfert et les services juridiques, pour voir sous quelle forme les étudiants peuvent y donner suite.

Mais j'ai tenu également à garder une ouverture sur un autre type de contenus, et j'aimerais le faire plus encore à l'avenir si les obligations de service le permettent. J'ai en effet pris en 2014 la **responsabilité** du cours-TD de **mathématique-informatique en L1 SFA**, qui est une option de choix large choisie cette année par environ 320 étudiants. Cela m'a permis d'une part d'intervenir auprès du public de Licence 1 (et de me familiariser avec les problématiques qui lui sont propres), et d'autre part d'enseigner des notions d'arithmétique et d'analyse de complexité très formatrices pour les étudiants. En outre, cela m'a permis d'avoir les bons interlocuteurs pour promouvoir avec eux le projet d'une licence mathématique-informatique, qui (je l'espère !) verra le jour en septembre prochain avec le principe suivant (proche de ce que j'avais eu la chance de faire dans le magistère ENS Cachan - Paris 7) : les étudiants valident en L2 et L3 40 ECTS dans chacune des deux licences, avec un choix d'UE guidé garantissant la compatibilité des emplois du temps, et ils obtiennent finalement les deux diplômes de licence. L'idée est bien entendu de prolonger le dispositif en Master par

la suite, afin de faire émerger quelques profils (les flux attendus ne dépassent pas la dizaine, mais ce serait déjà une réussite) présentant un niveau élevé dans et une double compétence - il se trouve déjà bon nombre de collègues désireux de proposer à de tels étudiants de poursuivre ensuite par un doctorat.

Depuis septembre 2016, j'ai pris la **Direction du Département de Mathématiques** (155 permanents, une cinquantaine de vacataires, plus de 20000h enseignées), passant donc la main sur la direction de la formation SID à ma collègue Cécile Chouquet. Cette responsabilité est très lourde, car le Département de Mathématiques est une composante de la Faculté des Sciences et d'Ingénierie qui est une structure relativement récente n'ayant pas encore trouvé de fonctionnement stable. Les principales tâches s'organisent autour de deux axes principaux : **en interne**, revoir et améliorer l'offre de formation à la lumière du contexte nouveau créé par la très grande difficulté financière de notre université. Les recrutements étant pour nous gelés pour plusieurs années, nous allons subir avec le départ de nos collègues à la retraite et en promotion une baisse significative de notre potentiel d'enseignement, tandis que des effectifs en hausse de 10% par an sont anticipés. En outre, l'offre de formation extrêmement éclatée à laquelle était arrivée le département présente de forts inconvénients, notamment en termes de lisibilité (par exemple, les étudiants n'ont pas moins que 5 voies possibles pour leur entrée en L1 de mathématiques) et par le caractère extrêmement lourd de sa gestion (pas moins de 35 personnes sont nécessaires pour assurer les diverses responsabilités d'année en mathématiques, ce qui explique une difficulté récurrente à les assumer correctement). **Sur un plan externe**, le principal travail consiste à construire avec les autres départements un mode de fonctionnement pour l'ensemble de la faculté, qui connaît de graves difficultés organisationnelles. Celles-ci ont culminé quelques semaines après ma prise de fonction par la grève de tous les secrétariats pédagogiques (une cinquantaine de personnes), après l'échec de la rentrée (où les difficultés ont été innombrables) et avant un deuxième échec prévisible au second semestre. Un travail de longue haleine est entamé, et la nouvelle équipe en responsabilité à la FSI s'attèle à assurer rapidement un retour à la normale, avant de pouvoir entreprendre des projets plus ambitieux notamment dans le cadre du PIA3. En outre, le département contribue en tant que composante de la FSI à la politique scientifique de l'université, en complémentarité du laboratoire (avec lequel nous travaillons en étroite collaboration). Je participe en particulier à la Commission Scientifique de la FSI, dont le rôle est notamment d'arbitrer les demandes de postes, les demandes de professeurs invités, etc.

Pour conclure cette présentation, je tiens à souligner que depuis mon recrutement à l'UPS je me suis **fortement impliqué** dans la promotion de nos formations, participant chaque année aux Journées Portes Ouvertes, au forum INFOSUP, allant dans des lycées et des IUT (en particulier mais pas seulement pour promouvoir la formation SID), etc.

## 2. Synthèse

### **Licence 1 :**

L1 SFA Choix Large Math-Info (CTD et TP), responsable CTD en 2014-2016

environ 300 étudiants par an, 20h de CTD + 16h de TP

Évolution en amphi-interactif / TD / TP et responsabilité en 2016-2017

plus de 300 étudiants, 12h d'amphi interactif, 18h de TD,

Encadrement du projet CMI de L1 (création et resp.) 2013-2016

une dizaine d'étudiants par an, 50h de travail personnel

**Licence 3 :**

L3 SID tutorat de stages

0 à 2 étudiants par an

**Master 1 :**

M1 SID Modèle linéaire généralisé (resp. cours et TD) 2013-2015

une trentaine d'étudiants par an

M1 MAPI3 Modélisation mathématique (création et resp. du cours) 2015-2016

30 étudiants

**Master 2 :**

M2P IMAT Data Mining (resp. cours, TD, TP) 2012-2016

entre 15 et 25 étudiants par an

M2P IMAT Méthodes parcimonieuses (resp. cours, TD, TP) 2012-2015

entre 15 et 25 étudiants par an

M2P SID Apprentissage Statistique (resp. cours, TD, TP, projets)

M2R statistique asymptotique (resp. cours)

M2R mathématiques du machine learning (co-resp. et création du cours)

M2 (mapi3 et SID) création avec l'INSA puis UT1, l'ENAC, Bordeaux, Pau, etc. du challenge (pour cette année voir <http://challenge.openbikes.co/>) M2P SID tutorat de stages

**Licence 3 - Master 1 - Master 2 :**

Projet de synthèse SID (création et resp. avec K. Pinel-Sauvagnat); encadrement du projet 2015 : "vizurbi".

**Encadrement de stages :**

- 2017 : Valentin Debarnot (stage fin d'étude INSA) : Optimisation du prix de réserve dans les enchères de second prix : apprentissage séquentiel (avec S. Gerchinovitz)
- 2016 : Timothée Mathieu (stage Master ENS Cachan) : Classification non supervisée pour les données massives (avec S. Gerchinovitz)
- 2016 : Florence Gourmelon (stage L3 SID) : site web de valorisation de l'étude des compétences au CIMI (avec C. Besse)
- 2015-2016 : Andressa Cerqueira (doctorat) : simulation parfaite des graphes aléatoires
- 2014 : Simon Coste (stage Master ENS Cachan) : algorithmes de bandits
- 2013 : Felix Bérourd (stage Licence ENS Lyon) : l'algorithme de Von Neumann-Peres pour la génération de bits aléatoires non biaisés
- 2013 : Isabelle Poilvet et Benoit Periol (stage L3 SID), création du site web de SID (avec K. Pinel-Sauvagnat)

### 3. Direction, animation et montage de formations

2016- : **Direction du département de mathématiques** (160 enseignants plus 92 intervenants externes, plus de 25000h d'enseignement)

2012- : **Responsable** de la **licence** et du **master** Statistique et Informatique Décisionnelle **SID - Big Data**

2013 : **création du CMI** en 2013

2015- : **Expert** pour le **réseau Figure** pour la création de nouveaux CMI, pour les évaluations à 2 ans

2014- : Représentant de l'UPS (au nom des CMI) pour le **groupe de travail** "événement phare" de **Toulouse Ingénierie**, participation à l'encadrement des 48h de l'innovation

2014- : **Responsable** de l'**UE de choix large Mathématiques-Informatique en L1-SFA** pour la partie cours-TD (cette année, 8 groupes de TD et environ 320 étudiants inscrits).

### 4. Rayonnement et activités internationales

2014 : **Cours** "machine learning et théorie des jeux" à l'**école d'été** pluridisciplinaire de théorie des jeux

## **Activité scientifique**

### 1. Thématiques de recherche

Mes travaux portent principalement sur la **modélisation stochastique et statistique**, et pour la plupart sur l'analyse et la création d'**algorithmes** d'estimation ou d'optimisation stochastique.

Ils sont marqués par mon premier sujet d'étude en doctorat : la **théorie de l'information**, et plus précisément le codage universel sans perte. Ce cadre convenait parfaitement à ma volonté de combiner mathématiques (probabilités et statistiques) et informatique (définition et analyse d'algorithmes) - volonté qui s'est confirmée et élargie par la suite. La théorie de l'information fournit un point d'entrée dans la modélisation stochastique idéal à bien des points de vue : d'abord, par la beauté d'une théorie où un problème parfaitement concret (la compression de données) a reçu à partir des travaux de Shannon un cadre mathématique rigoureux et riche, se prêtant à des travaux en interaction avec de nombreux champs des mathématiques (un exemple anecdotique : le résultat principal de mon article *A Lower-bound for the Maximin Redundancy in Pattern Processes* s'appuie essentiellement sur le dénombrement des partitions d'entiers à petits sommants, qui avait été faite par des techniques de combinatoire analytique). Ensuite parce que ce cadre est suffisamment riche et fort pour intégrer des approches très différentes : certains codeurs sont basées sur des idées très informatiques de grammaires, d'autres s'appuient explicitement sur des estimateurs - mais tous s'étudient et se comparent avec le même concept de redondance et les mêmes outils techniques. Enfin, parce que le modèle de Shannon, qui propose de considérer les messages à coder comme des réalisations de processus stochastiques, est suffisamment arbitraire pour qu'il impose une utilisation essentiellement instrumentale des modèles et conduise naturellement (avec une simplicité algébrique assez rare) à l'approche oracle qui est si féconde et pertinente en machine learning aujourd'hui.

C'est donc sans solution de continuité que mes travaux ont dérivé vers des questions de **sélection de modèles**, en particulier pour les **chaînes de Markov d'ordre variable**. Celles-ci ont été introduites par la théorie de l'information pour la modélisation parcimonieuse de sources à mémoire plus ou moins longue (en français par exemple, la loi de probabilité d'une lettre se situant après un symbole "espace" ne dépend guère de ce qu'il y avait avant, alors que pour prédire le symbole suivant un "u" il est très intéressant de savoir si l'on trouvait juste avant un "q" ou un "o"). Certains de ces travaux s'appuient sur des estimateurs dont l'inspiration est elle-même issue de la théorie du codage, et notamment du principe dit "Minimum Description Length" (**MDL**) selon lequel, entre plusieurs modèles décrivant les mêmes données, il faut choisir celui qui en donne la description la plus courte. Techniquement, j'ai compris progressivement qu'il fallait développer un certain outil pour obtenir les résultats (consistance, propriétés oracles) souhaités : des inégalités de déviation pour des sommes avec un nombre aléatoire de termes (très liées aux **processus auto-normalisés**). Je me suis également intéressé à d'autres modèles markoviens et en particulier aux **chaînes de Markov cachées**, sous l'angle à nouveau de l'estimation (avec des techniques semblables aux précédentes) mais aussi sous l'angle plus "traitement du signal" du **filtrage** (avec des méthodes **particulières**). Enfin, concernant les VLMC, je me suis également intéressé au problème de la simulation exacte de leur mesure stationnaire, en développant un algorithme de **couplage par le passé** tirant partie de l'idée de Propp et Wilson sans renoncer aux avantages de la parcimonie.

De façon assez surprenante, les inégalités de déviations auto-normalisées ont accompagné mon évolution thématique vers un sujet d'étude a priori assez distinct : **l'apprentissage par renforcement**. Après des premiers travaux sur les processus de décision markoviens partiellement observés (des modèles très généraux et très riches) je me suis rapidement recentré sur les **problèmes de bandits** (qui sont au contraire les plus simples), afin d'obtenir des résultats fins comblant l'écart qui existait entre bornes supérieures (garanties théoriques obtenues pour certains algorithmes performants) et bornes inférieures (limites intrinsèques indépassables des algorithmes passés et futurs). A partir de ces modèles, je me suis efforcé d'élargir ma culture concernant l'ensemble de **l'apprentissage statistique**, avec un intérêt particulier les méthodes **séquentielles**. Ces dernières (avec notamment les techniques de gradient stochastique) ont connu un très vif regain d'intérêt au cours des dix dernières années avec l'arrivée des big data et des flux de données ; elles ont constitué pour moi une porte d'entrée dans le monde de l'optimisation mathématique. Mon rôle d'animation (local et national) dans la recherche sur les Big Data m'a incité à apprendre les concepts fondamentaux à la base de la manipulation et de l'analyse des données (bases de données, intelligence artificielle), et à ouvrir mes travaux vers des champs applicatifs comme la gestion des contenus interactifs (projet ANR ALICIA, voir ci-dessous) ou la recherche d'information (avec notamment le co-encadrement d'une thèse avec deux collègues de l'IRIT). Parallèlement, je poursuis un certain nombre de projets autour des statistiques séquentielles dans le cadre du projet ANR SPADRO (avec Antoine Chambaz et Emilie Kaufmann notamment) et dans le cadre de co-encadrements de thèse, avec Fabrice Gamboa, Jean-Michel Loubes et Gilles Stoltz en particulier. Le trimestre thématique organisé en à l'automne 2015 a suscité, relancé ou accéléré ces projets, qui sont désormais un des axes forts de l'équipe projet AOC dont j'ai la responsabilité.

## 2. Publications

Je tente ici de décrire ma trajectoire scientifique à travers la très brève description de quelques articles.

Mes travaux sur la théorie de l'information ont porté sur différents sujets, mais l'aspect le plus abouti est certainement celui qui est développé dans l'article *Coding on countably infinite alphabets* publié en janvier 2009 dans les **IEEE Transactions on Information Theory** avec mes encadrants de thèse Stéphane Boucheron et Elisabeth Gassiat. Il développe une théorie générale du codage universel sans perte pour les sources sans mémoire sur les alphabets infinis, sous une hypothèse (incontournable) de majoration uniforme de ces sources par une enveloppe. Se mêlent donc les techniques classique théorie de l'information (le cas des alphabets finis étant bien connu) avec les outils modernes de statistique non non-paramétrique pour le réglage des équilibres biais-variance, l'obtention d'inégalités oracles et la détermination des taux de redondances minimax.

La théorie du codage universel a permis (par les travaux de J. Rissanen) l'introduction des modèles de Markov d'ordre variable (Variable Length Markov Chains, VLMC, également appelées sources à arbres de contexte). Une partie de mes travaux de thèse a consisté en l'étude statistique de ces modèles, et en particulier à l'estimation des modèles de mémoire. Ce n'est qu'une fois en poste à Telecom que j'ai vraiment développé le bon outil pour les étudier : des inégalités de déviation auto-normalisées (qui se sont d'ailleurs avérées essentielles pour l'étude des modèles de bandits évoqués plus loin). Cela m'a permis, avec ma collègue brésilienne Florencia Leonardi, d'aboutir dans l'article *Context Tree Selection : A Unifying View* publié Novembre 2011 dans **Stochastic Processes and their Applications** à une synthèse des différentes approches d'estimation et à une étude nouvelle de ce problème de sélection de modèles.

Mon intérêt s'est donc porté sur les modèles markoviens et leurs dérivés, comme les chaînes de Markov cachées à l'estimation desquelles d'autres de mes travaux. Les chaînes de Markov cachées constituaient une spécialité de l'équipe STA où j'ai passé cinq ans, et mes collègues Eric Moulines et Randal Douc m'ont associé à leurs travaux sur le filtrage particulière, une technique de traitement du signal nécessitant comportant à la fois un versant algorithmique intéressant et l'étude asymptotique de processus mélangeants. Est issu de ces recherches l'article *Sequential Monte Carlo smoothing for general state space Hidden Markov Models* publié en décembre 2011 aux **Annals of Applied Probability**.

Un autre aspect de mes travaux sur les modèles markoviens concerne la simulation exacte. Les chaînes de Markov d'ordre variable permettent une représentation parcimonieuse des modèles de mémoire, et les travaux statistiques mentionnés ci-dessus montrent qu'ils ne sont pas plus difficiles à manipuler et à mettre en œuvre que les chaînes de Markov classiques; toutefois, on pouvait se demander s'il en serait de même quant à la possibilité d'échantillonner sous leur mesure stationnaire : pour un résultat exact, perdrait-on la parcimonie et donc la possibilité de traiter des modèles à longue mémoire si l'on souhaitait appliquer une méthode de couplage par le passé comme celui de Propp et Wilson? L'article *Perfect Simulation Of Processes With Long Memory : A 'Coupling Into And From The Past' Algorithm*, publié dans **Random Structures and Algorithms** en mars 2015, montre que non. J'y décris un algorithme permettant, par un couplage original et par sa représentation soignée en mémoire, d'améliorer considérablement l'état de l'art tant sur le plan de l'efficacité pratique que concernant les conditions théoriques pour lesquelles la convergence de l'algorithme

est prouvée.

Parallèlement à ce sujet, mon intérêt s'est de plus en plus porté sur l'apprentissage statistique, en particulier sur l'apprentissage par renforcement ; après des premiers travaux sur des modèles très généraux de processus de décision markoviens partiellement observés, pour lesquels une analyse fine semble encore hors de portée, je me suis concentré sur des modèles plus simples mais avec la volonté de combler autant que possible l'écart entre théorie et pratique, entre algorithmes et analyse, entre bornes de regret prouvées et heuristiques utilisées dans les applications. Cet effort a donné naissance à une série d'articles dont un point de référence écrit en collaboration avec Olivier Cappé, Odalric-Ambrym Maillard, Rémi Munos, et Gilles Stoltz est intitulé *Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation*, publié aux **Annals of Statistics** en juin 2013. Nous y proposons la première analyse optimale des algorithmes de type Upper-Confidence Bound, à la fois dans un contexte paramétrique (familles exponentielles à un paramètre) et dans un contexte non-paramétrique (par des techniques d'estimation dites de vraisemblance empirique).

Terminons en mentionnant deux articles publiés dans le **Journal of Machine Learning Research**. Concernant le problème de l'identification des meilleurs bras dans les modèles de bandits, nous avons proposé avec Emilie Kaufmann et Olivier Cappé dans l'article *On the Complexity of Best Arm Identification in Multi-Armed Bandit Models* une étude des meilleurs algorithmes (bornes supérieures) et des limites théoriques (bornes inférieures) de ce qui peut être fait (un objectif qui sera pleinement accompli dans un article de 2016 avec Emilie Kaufmann, voir [20]). Enfin, en réponse à un besoin pratique d'analyse des systèmes électriques posé par Damien Ernst (il s'agissait de trouver une stratégie permettant d'utiliser le plus efficacement possible différentes heuristiques randomisées d'efficacité variables pour l'identification des configurations électriques dangereuses), j'ai proposé avec Sébastien Bubeck (Microsoft Research) de combiner l'heuristique optimiste d'UCB avec des estimateurs de masse manquante. Des garanties théoriques, ainsi qu'une analyse asymptotique de type limite fluide inhabituelle dans ce contexte, sont décrits dans *Optimal Discovery with Probabilistic Expert Advice : Finite Time Analysis and Macroscopic Optimality*, publié dans la même revue.

### 3. Encadrement doctoral et scientifique

**Direction de thèses :** 5 thèses soutenues, 5 thèses en cours (détails en annexe).

#### **Encadrement de stages :**

- 2017 : Valentin Debarnot (stage fin d'étude INSA et M2R) : Optimisation du prix de réserve dans les enchères de second prix : apprentissage séquentiel (avec S. Gerchinovitz)
- 2016 : Pierre Ménard (stage Master 2) : problèmes de bandits stochastiques et adversariaux (avec Gilles Stoltz)
- 2016 : Timothée Mathieu (stage Master ENS Cachan) : Classification non supervisée pour les données massives (avec S. Gerchinovitz)
- 2016 : Florence Gourmelon (stage L3 SID) : site web de valorisation de l'étude des compétences au CIMI (avec C. Besse)
- 2015-2016 : Andressa Cerqueira (doctorat) : simulation parfaite des graphes aléatoires

- 2014 : Simon Coste (stage Master ENS Cachan) : algorithmes de bandits
- 2013 : Felix Bérout (stage Licence ENS Lyon) : l'algorithme de Von Neumann-Peres pour la génération de bits aléatoires non biaisés
- 2013 : Isabelle Poilvet et Benoit Periol (stage L3 SID), création du site web de SID (avec K. Pinel-Sauvagnat)

#### 4. Diffusion et rayonnement

**Expertise.** Depuis octobre 2016, je suis **membre du Comité d'Évaluation Scientifique n° 40** "Mathématiques Informatique, Systèmes et Ingénierie de la communication". Le CES 40 est en charge des thématiques de mathématiques, mathématiques appliquées et informatique théorique, y compris désormais le traitement d'image et la robotique. Il s'agit d'opérer la sélection des dossiers financés (environ 160 demandes pour ce comité) en deux phases : une première phase où des pré-projets sont évalués uniquement par le comité (10-15 dossiers par membre du comité), puis une deuxième phase où les dossiers complets sont également étudiés par des experts désignés par le comité. Il s'agit d'un travail pluri-disciplinaire nécessitant (et renforçant) une vision assez large des mathématiques et d'une grande partie de l'informatique. J'y suis plus précisément chargé des projets relatifs aux probabilités et à la statistique en général, à l'apprentissage statistique et à la sciences des données en particulier. Par ailleurs, j'effectue de temps en temps des expertises pour divers organismes (comme le réseau Figure) ou prix (comme le prix Couseau l'an dernier).

#### **Activités éditoriales.**

**Associate Editor** du journal Stochastics depuis 2016.

**Éditeur invité** du journal ESAIM : Proceedings and Surveys vol. 51, Oct. 2015

**Membre du comité de programme** des JDS 2017, DSAA'2016, CAP'2016, BDMICS'16, ICML'15, et de la session Machine Learning de IJCAI 2015.

**Membre du comité scientifique** de JFPDA'14 et JFPDA'15.

Relecteur pour de nombreux journaux et conférences ( IEEE-Transaction on Information Theory, Annals of Statistics, Bernoulli, Journal of Machine Learning Research (JMLR), Mathematics of Operations Research, Stochastic Processes and their Applications, Polish journal Probability and Math Statistics, IEEE-Transaction on Signal Processing, IEEE Transactions on Automatic Control, AStA Advances in Statistical Analysis, Communications in Statistics - Theory and Methods, Automatica, conférences NIPS, COLT, ALT, ICML, ISIT, Eusipco, etc.).

#### **Participation à 7 jurys de thèses (dont 1 rapporteur) et 3 HDR (dont 1 rapporteur) :**

- thèse Sandro Gallo (2010, USP-Brésil),
- thèse de Léa Laporte (2013, Paul Sabatier-IRIT)
- thèse de Guillaume Mijoule (2013, Paul-Sabatier IMT)
- HDR Sébastien Loustau (2014, Angers)
- (président) thèse de Benjamin Girault (2015, ENS Lyon)
- (président) HDR de Matthieu Serrurier (2015, Paul Sabatier-IRIT)
- (rapporteur) thèse Jérémy Bensadon (2016, Orsay-LRI)

- (rapporteur & président) HDR de Michal Valko (2016, ENS Cachan)
- thèse d'Émile Contal (2016, ENS Cachan)
- (président) thèse de Sofiane Saadane (2016, Paul-Sabatier IMT)
- (président) HDR Nicolas Couellan (2016, Paul-Sabatier IMT)

### **Diffusion du savoir.**

J'ai donné diverses conférences grand public sur les big data en diverses occasions (j'ai notamment donné une conférence à l'IREM de Boreaux en janvier 2017, et participé à une table ronde pour les 20 ans de l'ENSAI en octobre 2016).

J'ai également donné quelques interviews pour des journalistes, dont une rencontre "petit déjeuner du CNRS" organisée par l'AJSPI (Association des Journalistes Scientifiques de la Presse d'Information) le 21 Novembre 2013 sur le thème des Big Data. Cela a donné suite à divers articles (Le Monde, La Libre Belgique, L'Usine Digitale, Toulemploi...).

Je suis également **membre du bureau de groupe MAS de la SMAI** depuis août 2012. Le groupe thématique Modélisation Aléatoire et Statistique de la Société de Mathématiques Appliquées et Industrielles, a pour but de promouvoir auprès du monde académique ainsi que des entreprises les méthodes de la statistique et des probabilités appliquées. Dans cet objectif, le groupe thématique MAS organise des activités scientifiques et techniques : Journées MAS, Journées Thématiques, Colloques, Prix de thèse Jacques Neveu, etc.

**Organisation des Journées MAS 2014.** J'ai été responsable de l'organisation des Journées MAS 2014 à Toulouse. Cette rencontre, organisée tous les deux ans par le groupe MAS de la SMAI, réunit probabilistes et statisticiens autour d'une thématique principale, avec des sessions thématiques couvrant un large spectre ; elle est aussi l'occasion de remettre le prix de thèse Jacques Neveu. L'édition toulousaine de 2014 a réuni plus de 200 participants, avec pour thème les *phénomènes de grande dimension*. Elle a innové en introduisant notamment une session poster pour les étudiants, qui été dotée d'un prix sponsorisé par l'éditeur Springer.

**Organisation des Journées Big Data 2013-2017.** Chacune de ces trois dernières années, j'ai co-organisé avec mes collègues (principalement Philippe Besse, Jean-Michel Loubès et Nathalie Aussenac-Gilles) des journées consacrées aux Big Data. Ces journées ont été conçues autour du triangle "formation-recherche-industrie" qui est au cœur du projet des Coursus de Master en Ingénierie (voir ) Elles associaient donc les laboratoires IMT, IRIT et CIMI, les formations toulousaines concernées (notamment le CMI SID, l'INSA, puis IMAT, TSE et d'autres écoles), et les partenaires industriels tant des premiers (à travers les projets de recherche communs) que des seconds (avec notamment les membres du comité de perfectionnement du CMI SID, comme par exemple Laurent Peltiers venu présenter les problématiques données des essais en vols chez Airbus). Grâce à Fabrice Gamboa, l'édition 2015 a combiné deux événements : une première journée plus particulièrement orientée vers la recherche, co-organisée par les GDR MaDICS et Mascotnum du CNRS ; et une deuxième journée, au format assez semblable aux sessions précédente. En 2016-2017, une journée semblable dans la lignée des précédentes a été organisée à l'université Toulouse 1 (qui participait à nos journées depuis le début), initiant une rotation entre les sites toulousains impliqués. Elle s'est terminée par la remise des prix du challenge opendata organisé pour nos

différentes formations, élargi cette année à Bordeaux, Pau (avec la participation d'une équipe de l'Ecole Polytechnique) : voir <http://challenge.openbikes.co/>.

**Organisation du trimestre thématique CIMI Machine Learning.** Le but de ce trimestre thématique, que j'ai porté avec des collègues mathématiciens et informaticiens, était de proposer une série d'événements scientifiques reflétant les intérêts communs, actuels et potentiels, de l'IMT et de l'IRIT en Machine Learning. Il s'inscrivait dans la continuité d'un groupe de travail (co-organisé avec Sébastien Gerchinovitz) sur cette thématique, qui regroupe des collègues des deux disciplines et de plusieurs laboratoires toulousains. Son déroulement s'est étalé sur les mois de septembre à décembre 2015. L'accent s'est porté sur trois axes : l'optimisation pour l'apprentissage et le traitement d'image, l'apprentissage sur les données structurées et le traitement des langages naturels, et l'apprentissage séquentiel. Certains de ces événements étaient en partie dédiés à étudiants de master, de doctorat ou de post-doctorat. Outre les invitations de spécialistes internationaux (Ronen Eldan, Sébastien Bubeck, Johan Suykens, Peter Richtárik, etc.), ses principaux événements marquants étaient une école d'été, trois workshops et deux journées thématiques. Les débouchés de cet événement sont encore difficiles à mesurer en termes quantitatifs (publications, encadrements de thèses, etc.). On peut toutefois déjà souligner son apport structurant pour nos laboratoires : en résulte notamment actuellement le montage d'une équipe-projet rassemblant des collègues de plusieurs équipes de l'IMT et de l'IRIT souhaitant travailler ensemble en bousculant les frontières disciplinaires et les habitudes (voir item suivant).

**Participation à un réseau de recherche : directeur adjoint du GDR MaDICS.** Le GDR MaDICS (CNRS-GDR3708) est une création récente du CNRS. Il est né officiellement le premier janvier 2015, après une phase de montage en 2013-2014 à laquelle j'ai été associé. J'y représente la section 41 (co-tutelle du GDR avec les deux sections informatique), et thématiquement l'analyse statistique de données ainsi que l'apprentissage machine. Notre travail est notamment de mobiliser les communautés concernées à dépasser les frontières disciplinaires et communautaires pour s'engager dans des projets transverses, avec les producteurs de données, qui auraient du mal à émerger sans stimulation. Après une assemblée générale (ou j'ai tenté de faire en sorte que les défis concernant l'analyse de données soient bien présentés), nous organisons à présent des appels à projets pour encourager des actions allant en ce sens. Afin de construire un dialogue constructif avec des communautés parfois très éloignées de celles que je côtoie habituellement, ce travail très pluridisciplinaire m'encourage à élargir mes domaines de compétences et à développer une vision plus large de la recherche et des laboratoires.

## 5. Responsabilités scientifiques

**Création puis direction de l'équipe-projet AOC.** A la suite du groupe de travail sur le machine learning et du trimestre thématique, et dans le cadre du projet du Labex CIMI, j'ai porté le montage d'une structure nouvelle permettant aux chercheurs intéressés (qu'ils soient à l'IMT, dans l'équipe de probabilité/statistique ou dans l'équipe d'analyse appliquée, ou bien à l'IRIT dans l'équipe d'intelligence artificielle, de traitement du langage, ou de recherche d'information par exemple) de se regrouper dans les mêmes locaux au sein d'une structure intégrée afin de développer activement

des collaborations, sans renoncer à l'appartenance à leur équipe d'origine. L'équipe-projet, qui comporte une vingtaine de permanents, a commencé ses travaux en septembre 2016, et j'en ai pris la responsabilité. Elle s'organise autour de différents projets de recherche, co-encadrements, et l'organisation d'événements (pour cette année, les rencontres franco-chiliennes d'optimisation organisées par Edouard Pauwels par exemple).

**Porteur du projet ANR blanc SPADRO** (Statistique Semi-Paramétrique pour l'Allocation Dynamique de Ressources et l'Optimisation), programme blanc 2013, avec Antoine Chambaz (Université Paris 10 Nanterre). Ce projet entend proposer de nouvelles méthodes et de nouvelles analyses pour les problèmes d'allocation dynamique de ressources, en provoquant un brassage fécond de la théorie de l'apprentissage avec la statistique semi-paramétrique.

**Responsable du partenaire UPS/IMT du projet ANR ALICIA** (Apprentissage Adaptatif pour le Crowdsourcing Intelligent et l'Accès à l'Information), du programme Contents numériques et interactions (CONTINT) 2013. Ce projet vise à contribuer au développement de mécanismes d'apprentissage fortement adaptatifs pour des sources et besoins en information dynamiques et fortement contextualisés – des composants clés des applications sociales et centrées sur les utilisateurs – tout en promouvant la pertinence de l'information, la complétude et la diversité de la manière dont les contenus et les utilisateurs sont sélectionnés en réponse aux besoins. Pour atteindre ce but, nous nous concentrons sur des algorithmes d'apprentissage adaptatif qui ont le potentiel de bien se comporter dans les conditions des environnements en ligne, tels que la famille des bandits manchots (multi-armed bandits).

**Projet industriel.** Dans un cadre initié par un projet étudiant, réalisation avec la société Inbenta d'un outil innovant de segmentation automatique de FAQ tirant partie des connaissances linguistiques de la société (basées sur la théorie sens-texte) et des d'approches récentes du clustering basée sur l'échantillonnage intelligent.

**Cours d'école d'été.** 3 séances de 2h sur le thème "machine learning et théorie des jeux" à l'école d'été pluridisciplinaire de théorie des jeux d'Aussois, en septembre 2014.

**MaDICS et MDK.** A côté de la direction adjointe du GDR MaDICS, je suis membre du projet ANR réseau de recherche MDK : Réseau Données Massives et Connaissances. Il a pour objectif de donner un cadre à une réflexion collective, par l'ensemble de ces communautés, sur la recherche et l'innovation dans le domaine des Masses de Données et Connaissances. Il s'agit de produire des études prospectives, de donner des recommandations et de proposer des Actions concrètes. Cette réflexion pourra s'accompagner d'une stratégie tant sur le plan national qu'europpéen.

**Partenariat avec le Brésil.** Après quatre séjour à l'Universidad de Sao Paulo au Brésil, liens de recherche que j'y ai tissés ont permis la venue pour l'année 2015-2016 d'Andressa Cerqueira pour passer un an de doctorat sous ma direction à l'IMT.

**Expert** pour divers appels d'offre (dont le CES 40 de l'ANR), **rapporteur** pour le prix de thèse Paul Caseau de l'Académie des technologies et EDF, etc.

## **Responsabilités collectives**

### *1. Présentation générale des responsabilités particulières*

Comme je l'ai déjà développé dans les sections antérieures, mon profil de mathématicien ayant une culture informatique et travaillant sur des thématiques (machine learning) parfois classées à l'étranger en 'computer science' me confère un rôle un peu particulier au sein de l'IMT et font souvent de moi l'interlocuteur privilégié dans les relations avec l'informatique. Je ne reviens pas sur la **direction adjointe du GDR MADICS**, ni sur la **direction du département de mathématiques**, ni sur la **responsabilité du CMI SID** et la participation à la commission événement-phare de Toulouse-Ingénierie, ni sur le montage du **trimestre thématique** sur le machine learning, ni sur la **responsabilité de l'équipe-projet AOC**, tous évoqués plus haut. Jusqu'à ma prise de fonction à la tête du département j'étais également membre du **comité exécutif du Labex CIMI** (Centre International de Mathématiques et d'Informatique de Toulouse), qui gère plusieurs appels à projets concernant principalement les personnels de l'IMT et de l'IRIT, ainsi que des bourses doctorales, post-doctorales, etc., et qui cherche à encourager les interfaces entre mathématique et informatique. J'essaye de travailler à ce que l'**organisation des journées Big Data** se fasse en bonne intelligence entre les deux laboratoires, et avec d'autres établissements toulousains. Je suis membre du **bureau du conseil scientifique** de l'IMT, et j'ai été sollicité pour **présider les comité de sélection** qui ont recruté en 2015 un maître de conférence sur un poste "croisé" IMT-IRIT (le poste était ouvert aux candidats des sections 26 et 27 ; le jury était paritaire ; un candidat relevant de la section 27 aurait été affecté à l'IMT, et le candidat retenu, provenant des mathématiques, a été affecté à l'IRIT) et en 2016 un maître de conférence en probabilité et statistique. Je suis membre du **groupe d'avancement** rang A de l'IMT, qui pour ces questions est regroupé avec l'IRIT au sein du même pôle. En 2016 j'ai intégré le **comité de prospective de l'IMT**, qui réfléchit aux positionnement scientifique des mathématiques toulousaines et propose des directions de développement. J'ai également intégré le **CES 40 de l'ANR**, comité d'évaluation scientifique des dossiers en mathématiques, informatique théorique et une partie du traitement du signal.

### *2. Responsabilités administratives*

2016- : **Directeur du département de mathématiques**

2016- : **Responsable de l'équipe-projet AOC**

2016-2017 : Membre du **comité de prospective de l'IMT**

2014-2016 : Membre du **comité exécutif du Labex CIMI**

2013- : Montage puis **direction adjointe du GDR MaDICS**

2013- : Membre du **bureau du conseil scientifique de l'IMT**

2013- : Membre du **groupe d'avancement rang A de l'IMT**

2014- : **Vice-président** et deux fois **Président de comités de sélection**

### 3. Responsabilités dans les projets et la vie collective de l'établissement

2013- : représentant de l'UPS dans la commission "événement phare" de Toulouse Ingénierie

### 4. Responsabilités et mandats nationaux, ou régionaux

2009-2012 : **responsable du jury de mathématique du concours d'entrée à l'ENS, voie BL**

Expert pour divers appels d'offres

## **Annexes**

### Liste classée des publications

Conformément aux habitudes des communautés auxquelles ils se rattachent, certains articles sont publiés dans des actes de conférence sans faire l'objet de publication ultérieure dans des journaux ; en effet, des conférences comme COLT, NIPS (et dans une moindre mesure Allerton ou ALT) consistent des références dans lesquelles une publication est aussi sélective et reconnue qu'une revue internationale.

### **Ouvrages collectifs et individuels**

1. *Les systèmes de recommandations en ligne*, dans *Les Big Data à Découvert*  
Ouvrage collectif  
**Editions CNRS**
2. *Modélisation Aléatoire et Statistique - Journées MAS 2014*  
Editeur invité : Aurélien Garivier  
**ESAIM : Proceedings and Surveys vol. 51 Oct. 2015**
3. *Exercices d'oral de mathématiques BL - ECE - ECS corrigés et commentés par leurs auteurs*  
Sylvain Arlot, Aurélien Garivier et Gilles Stoltz  
**Editions Ellipses collection Références Sciences, 2015.**

### **Articles publiés dans des revues internationales à comité de lecture**

4. *Bregman superquantiles. Estimation methods and applications*  
Tatiana Labopin-Richard, Fabrice Gamboa, Aurélien Garivier, Bertrand Iooss  
**Dependence Modeling vol. 4(1), Mar. 2016, pp.76-108**
5. *On the Complexity of Best Arm Identification in Multi-Armed Bandit Models*  
Emilie Kaufmann, Olivier Cappé, Aurélien Garivier  
**Journal of Machine Learning Research vol. 17 (1) Jan. 2016 pp.1-42**
6. *Perfect Simulation Of Processes With Long Memory : A 'Coupling Into And From The Past' Algorithm*  
Aurélien Garivier  
**Random Structures and Algorithms vol. 46 (2) Mar. 2015 pp.300-319**

7. *Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation*  
Olivier Cappé, Aurélien Garivier, Odalric-Ambrym Maillard, Rémi Munos, Gilles Stoltz  
**Annals of Statistics vol. 41 (3) Jun. 2013 pp.1516-1541**
8. *Joint Estimation of Intersecting Context Tree Models*  
Antonio Galves, Aurélien Garivier, Elisabeth Gassiat  
**Scandinavian Journal of Statistics vol. 40 (2) Jun. 2013 pp.344-362**
9. *Optimal Discovery with Probabilistic Expert Advice : Finite Time Analysis and Macroscopic Optimality*  
Sébastien Bubeck, Damien Ernst, Aurélien Garivier  
**Journal of Machine Learning Research vol. 14 Feb. 2013 pp.601-623**
10. *Sequential Monte Carlo smoothing for general state space Hidden Markov Models*  
Randal Douc, Aurélien Garivier, Eric Moulines, Jimmy Olsson  
**Annals of Applied Probability vol. 21 (6) Dec. 2011 pp.2109-2145**
11. *Context Tree Selection : A Unifying View*  
Aurélien Garivier, Florencia Leonardi  
**Stochastic Processes and their Applications vol. 121 (11) Nov. 2011 pp.2488-2506**
12. *Optimally Sensing a Single Channel Without Prior Information : The Tiling Algorithm and Regret Bounds*  
Sarah Filippi, Olivier Cappé, Aurélien Garivier  
**IEEE Journal of Selected Topics in Signal Processing vol. 5 (1) Feb. 2011 pp.68-76**
13. *On Approximate Maximum Likelihood Methods for Blind Identification : How to Cope with the Curse of Dimensionality*  
Steffen Barambruch, Aurélien Garivier, Eric Moulines  
**IEEE Transactions on Signal Processing vol. 57 (11) Nov. 2009 pp.4247-4259**
14. *A Lower-bound for the Maximin Redundancy in Pattern Processes*  
Aurélien Garivier  
**Entropy vol. 11 (4) Oct. 2009 pp.634-642**
15. *A MDL approach to HMM with Poisson and Gaussian emissions. Application to order identification* Antoine Chambaz, Aurélien Garivier, Elisabeth Gassiat  
**Journal of Statistical Planning and Inference vol. 139 (3) Mar. 2009 pp.962-977**
16. *Coding on countably infinite alphabets*  
Stéphane Boucheron, Aurélien Garivier, Elisabeth Gassiat  
**IEEE Transactions on Information Theory vol. 55 (1) Jan. 2009 pp.358-373**
17. *Redundancy of the Context-Tree Weighting Method on Renewal and Markov Renewal Processes*  
Aurélien Garivier  
**IEEE Transactions on Information Theory vol. 52 (12) Dec. 2006 pp.5579-5586**
18. *Consistency of the unlimited BIC Context Tree estimator*  
Aurélien Garivier  
**IEEE Transactions on Information Theory vol. 52 (10) Oct. 2006 pp.4630-4635**

**Articles publiés dans des actes de conférences internationales à comité de lecture**

19. *On Explore-Then-Commit Strategies*  
Aurélien Garivier, Emilie Kaufmann, Tor Lattimore  
**Neural Information Processing Systems 30 Dec. 2016, Proceedings published as Advances in Neural Information Processing Systems 29**
20. *Optimal Best Arm Identification with Fixed Confidence*  
Aurélien Garivier, Emilie Kaufmann  
**Conference On Learning Theory Jun. 2016**
21. *Maximin Action Identification : A New Bandit Framework for Games*  
Aurélien Garivier, Emilie Kaufmann, Wouter Koolen  
**Conference On Learning Theory Jun. 2016**
22. *A Multiple-play Bandit Algorithm Applies to Recommender Systems* Jonathan Louëdec, Max Chevalier, Josiane Mothe, Aurélien Garivier, Sébastien Gerchinovitz  
**Florida Artificial Intelligence Research Society Conference 28 May. 2015 pp.67-72**
23. *On the Complexity of A/B Testing* Emilie Kaufmann, Olivier Cappé, Aurélien Garivier  
**Conference On Learning Theory Jun. 2014 pp.461-481**
24. *Informational Confidence Bounds for Self-Normalized Averages and Applications*  
Aurélien Garivier  
**IEEE Information Theory Workshop Sep. 2013 pp.489-493**
25. *Optimal Discovery with Probabilistic Expert Advice*  
Sébastien Bubeck, Damien Ernst, Aurélien Garivier  
**IEEE Conference on Decision and Control 51 Dec. 2012 pp.6808-6812**
26. *On Bayesian Upper Confidence Bounds for Bandit Problems*  
Emilie Kaufmann, Olivier Cappé, Aurélien Garivier  
**International Conference on Artificial Intelligence and Statistics 15 Apr. 2012 pp.592-600**
27. *On Upper-Confidence Bound Policies for Non-stationary Bandit Problems*  
Aurélien Garivier, Eric Moulines  
**Algorithmic Learning Theory 22 Oct. 2011 pp.174-188 Proceedings published as Lecture Notes in Computer Science Volume 6925**
28. *The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond*  
Aurélien Garivier, Olivier Cappé  
**Conference On Learning Theory 24 Jul. 2011 pp.359-376**
29. *Parametric Bandits : The Generalized Linear Case*  
Sarah Filippi, Olivier Cappé, Aurélien Garivier, Csaba Szepesvári  
**Neural Information Processing Systems 24 Dec. 2010 Proceedings published as Advances in Neural Information Processing Systems 23 (with supplementary material)**
30. *Optimism in Reinforcement Learning and Kullback-Leibler Divergence*  
Sarah Filippi, Olivier Cappé, Aurélien Garivier  
**Allerton Conference on Communication, Control, and Computing 48 Sep. 2010 pp.115-122**

31. *Pseudo Regenerative Block-Bootstrap for Hidden Markov Models*  
Stéphan Cléménçon, Aurélien Garivier, Jessica Tressou  
**IEEE Statistical Signal Processing Workshop 15 Aug. 2009 pp.465-468**
32. *On Approximate Maximum Likelihood Methods for Blind Identification : How to Cope with the Curse of Dimensionality*  
Steffen Barenbruch, Aurélien Garivier, Eric Moulines  
**International Workshop on Signal Processing Advances for Wireless Communications 9 Jun. 2008 pp.639-643**
33. *On Optimal Sampling for Particle Filtering in Digital Communication*  
Steffen Barenbruch, Aurélien Garivier, Eric Moulines  
**International Workshop on Signal Processing Advances for Wireless Communications 9 Jun. 2008 pp.634-638**
34. *An event spacing experiment*  
Aurélien Garivier, Anthony Winstanley, Mark R. Greenstreet  
**International Symposium on Asynchronous Circuits and Systems 8 Apr. 2002 pp.47-56**

#### **Articles publiés dans des revues nationales à comité de lecture**

35. *Algorithmes de bandit pour la recommandation à tirages multiples*  
Jonathan Louëdec, Max Chevalier, Aurélien Garivier, Josiane Mothe  
**Document numérique (Hermès) vol. 18 :2-3 2015 pp.59-79**
36. *Big Data - Retour vers le Futur 3 ; De Statisticien à Data Scientist*  
Philippe Besse, Aurélien Garivier, Jean-Michel Loubès  
**Ingénierie des Systèmes d'Information vol. 19 (3) Mar. 2014 pp.93-105**

#### **Articles publiés dans dans des actes de conférences nationales à comité de lecture**

37. *Systèmes de recommandations : algorithmes de bandits et évaluation expérimentale*  
Jonathan Louëdec, Max Chevalier, Aurélien Garivier, Josiane Mothe  
**Journées de statistiques 2015 Jun. 2015**
38. *Algorithmes de bandit pour les systèmes de recommandation : le cas de multiples recommandations simultanées*  
Jonathan Louëdec, Max Chevalier, Aurélien Garivier, Josiane Mothe  
**Conférence en Recherche d'Information et Applications CoRIA Mar. 2015 pp.73-88**
39. *Poursuite d'étude après un IUT STID : l'exemple du Cours de Master en Ingénierie Statistique et Informatique Décisionnelle de Toulouse*  
Cécile Chouquet, Aurélien Garivier  
**Journées de statistiques 2014 Jun. 2014**
40. *Bandits contextuels : apprentissage par renforcement dans les modèles linéaires généralisés*  
Sarah Filippi, Olivier Cappé, Aurélien Garivier, Csaba Szepesvári  
**Journées Francophones Planification, Décision, et Apprentissage pour la conduite de systèmes Jun. 2010**

## **Autres publications**

41. Analyse d'algorithmes pour l'estimation et l'optimisation stochastiques  
Aurélien Garivier  
**Mémoire d'habilitation à diriger des recherches Nov. 2011**
42. *Modèles contextuels et alphabets infinis en théorie de l'information*  
Aurélien Garivier  
**Mémoire de thèse Nov. 2006**
43. *Parsing et codage universel*  
Aurélien Garivier  
**Rapport de DEA Sep. 2003**

## **Pré-publications (soumis, disponibles sur hal/Arxiv)**

44. *Learning the distribution with largest mean : two bandit frameworks*  
Emilie Kaufmann, Aurélien Garivier
45. *Explore First, Exploit Next : The True Shape of Regret in Bandit Problems*  
Aurélien Garivier, Pierre Ménard, Gilles Stoltz
46. *Conditional quantile sequential estimation for stochastic codes*  
Tatiana Labopin-Richard, Fabrice Gamboa, Aurélien Garivier
47. *Oracle approach in context tree estimation*  
Matthieu Lerasle, Aurélien Garivier

## **Principales conférences invitées**

1. Aug. 2016 - *Journées MAS 2016, Grenoble* - Sur la complexité de l'identification du meilleur bras sous contrainte de risque dans un modèle de bandits
2. Jan. 2016 - *Multi-armed Bandit Workshop 2016 at STOR-i, Lancaster University, UK* - Optimal Discovery with Probabilistic Expert Advice.
3. Mar. 2015 - *Information Theory, Learning and Big Data Workshop, Simons Institute, Berkeley, USA* - On the Complexity of Best Arm Identification in Multi-Armed Bandit Models.
4. Oct. 2014 - *London Probability Seminar : Multi-armed bandits meeting, Imperial College, London, UK* - Bandits for Exploration : Best Arm Identification and Discovery with Probabilistic Experts.
5. Sep. 2014 - *Ecole d'été pluridisciplinaire de Théorie des Jeux, Aussois* - Machine Learning
6. Jul. 2014 - *Conference on Stochastic Processes and their Applications, SPA'14, Buenos Aires, Argentina* - Perfect Simulation of Processes With Long Memory : A "Coupling Into and From The Past" Algorithm.
7. Jun. 2014 - *Journées Statistiques du Sud, Barcelona, Spain* - Empirical Likelihood Upper Confidence Bounds For Bandit Models.
8. Apr. 2014 - *StatLearn 2014, AgroParisTech, Paris* - Optimistic Solutions for Dynamic Resource Allocation.
9. Dec. 2013 - *Rencontres de Statistique Mathématique "Mathematical Statistics with Applications in Mind", CIRM, Marseille* - Optimal Discovery with Probabilistic Expert Advice : Finite Time Analysis and Macroscopic Optimality

10. Sep. 2013 - *IEEE Information Theory Workshop*, **Sevilla, Spain** - Informational Confidence Bounds for Self-Normalized Averages and Applications.
11. Mar. 2013 - Fête Parisienne in Computation, Inference and Optimization, **IHES, Bures sur Yvette** - Dynamic resource allocation as an estimation problem.
12. Jan. 2012 - *Séminaire parisien*, **Paris** - Apprentissage par renforcement et déviations auto-normalisées.
13. Jul. 2011 - *Jorma's Razor 2, conférence en l'honneur de Jorma Rissanen*, **Sao Sebastiao, Brésil** - A Propp-Wilson algorithm for Context Tree Sources.
14. Sep. 2010 - *Journées MAS 2010*, **Toulouse** - Arbres de Contextes Probabilisés.
15. Aug. 2009 - *Ecole Brésilienne de Probabilités de UPS*, **Maresias, Brésil** - Context tree models for source coding.
16. Jul. 2009 - *European Meetings of Statisticians*, **Toulouse** - Coding on infinite alphabets.

Viennent bien sûr s'y ajouter nombre d'exposés dans différents laboratoires, en particulier français (Paris 7, Nanterre, CMAP, Mascot-num, JDS, CIMI, TSE, INRA, LPT, LAAS, Telecom ParisTech, Montpellier, Engref, SMILE, Sao Paulo, Ottawa, etc.).

### **Brevets, licences, logiciels**

- **PyMaBandits**, framework d'expérimentation matlab et python accessible sur <http://mloss.org/software/view/415/>
- Package matlab pour la simulation exacte de processus markoviens et de VLMC, accessible sur ma page web
- Petite étude comparative python / R / scilab / octave / matlab / C++ / C sur l'algorithme de Baum-Welch, accessible sur ma page web

### Directions de thèses

5 thèses soutenues, 5 thèses en cours.

### **Thèses soutenues**

**Sarah Filippi** thèse avec Orange Labs soutenue le 24 novembre 2010. Sujet : sur les stratégies optimistes en apprentissage par renforcement, à 50% avec Olivier Cappé (Telecom ParisTech). 4 publications pendant la thèse.

**Emilie Kaufmann** soutenue le premier octobre 2014. Sujet : analyse de stratégies bayésiennes et fréquentistes pour l'allocation séquentielle de ressources à 33% avec Olivier Cappé (Telecom ParisTech) et Rémi Munos (INRIA Lille). **Prix de thèse Jacques Neveu 2014**. 5 publications pendant la thèse.

**Tatiana Labopin-Richard** soutenue le 1er juillet 2016. Sujet : en thèse sur les méthodes statistiques et d'apprentissage pour la validation d'un code numérique stochastique (application aux énergies renouvelables), à 50% avec Fabrice Gamboa (IMT). 3 publications issues de la thèse.

**Wenjuan Wang** soutenue le 16 septembre 2016. Sujet : optimisation pour l'apprentissage statistique de données incertaines, à 33% avec Nicolas Couellan (IMT) et Kers-  
tin Bystricky (Laboratoire de Biologie Moléculaire Eucaryote). 3 publications pendant  
la thèse.

**Jonathan Louédéc** soutenue le 4 novembre 2016. Sujet : Stratégies de Bandit pour  
les Systèmes de Recommandation, à 33% avec Josiane Mothe et Max Chevalier (IRIT).  
4 publications pendant la thèse.

### Liste des thèses en cours

**Laurent Rossi** (professeur en lycée) en thèse sur la statistique semi-paramétrique  
pour l'allocation dynamique de ressources et l'optimisation.

**Grégoire Corbière** en thèse sur la détection de comportements déviants dans un en-  
vironnement de type cloud, à 33% avec Jean-Michel Loubes (IMT) et Mathieu Serrurier  
(IRIT).

**Pierre Ménard** en thèse sur les bandits stochastiques et plans d'expériences adap-  
tatifs (contribution à une approche statistique non-paramétrique), à 50% avec Gilles  
Stoltz (CNRS, HEC Paris)

**Mastane Achab** en thèse sur l'apprentissage statistique par renforcement et les  
risques extrêmes, à 33% avec Stéphan Cléménçon et Anne Sabourin (Telecom Pa-  
risTech).

**Léonard Torossian** en thèse sur la métamodélisation et l'optimisation robuste – ap-  
plication à la recherche d'idéotypes sous incertitude climatique, à 33% avec Robert  
Faivre et Victor Picheny (INRA Toulouse)..

Cyrille Laroche, qui avait entamé une thèse sur le développement d'outils pour  
l'analyse de données multidimensionnelles en vue d'une aide à la décision en temps  
réel, à 33% avec Jean-Michel Loubes (IMT) et le Marine Depecker (CEA Saclay), a été  
embauché avant la fin de sa thèse.

### Devenir des docteurs

Sarah Filippi est Lecturer au Department of Statistics de l'université d'Oxford.

Emilie Kaufmann est chargée de recherche au CNRS au laboratoire CRISTaL de Lille.  
Elle avait reçu pour sa thèse le Prix Jacques Neveu.

Tatiana Labopin-Richard a pris un poste d'enseignante agrégée dans l'enseignement  
secondaire.

Wenjuan Wang est Post-doc à l'université d'Oxford (Angleterre).

Jonathan Louédéc est data scientist chez CDiscount France, où il continue à avoir  
une activité de recherche.