

1. PARCOURS PROFESSIONNEL

Maîtresse de conférences <i>Laboratoire Paul Painlevé</i>	Université de Lille, France <i>Depuis September 2017</i>
○ 02/22 – 12/22 : délégation CNRS ○ 04/18 – 08/18 et 04/21 – 08/21 : congés maternités	
Postdoctorante <i>Laboratoire MICS</i>	CentraleSupélec, France <i>Octobre 2016 – Août 2017</i>
Postdoctorante <i>Centre for Environmental and Climate Research</i>	Lund University, Suède <i>Juin 2015 – Septembre 2016</i>
Postdoctorante <i>Institut de Modélisation des Systèmes Vivants</i>	Université Paris-Saclay, France <i>Avril 2014 – Mai 2015</i>
ATER <i>Laboratoire MICS</i>	CentraleSupélec, France <i>Septembre 2013 – Février 2014</i>

2. FORMATION

Doctorat de Mathématiques Appliquées <i>CentraleSupélec</i>	Antony <i>2010 – 2014</i>
M2 Recherche de Statistiques <i>Sorbonne Université</i>	Paris <i>2007–2008</i>
ISUP (2ème et 3ème années) <i>Sorbonne Université</i>	Paris <i>2006–2008</i>
M1 de Mathématiques <i>Sorbonne Université</i>	Paris <i>2005–2006</i>
Licence de Mathématiques <i>Sorbonne Université</i>	Paris <i>2002–2005</i>

3. TRAVAUX DE RECHERCHE

3.1. Activités de recherche.....

Mes travaux de recherche se situent à l'interface entre les statistiques computationnelles et les statistiques appliquées. Je travaille principalement sur les modèles à variables latentes, avec des applications en

biologie et en écologie.

Modèles non-linéaires à effets mixtes

Une grande partie de mes activités de recherche porte sur les modèles non linéaires à effets mixtes (MNLM), même si certains des résultats développés dans ce cadre peuvent également être adaptés à des modèles à variables latentes plus généraux. Je détaille ci-dessous les deux principaux problèmes qui m'ont intéressé : l'estimation dans les NLMM et les tests d'hypothèses.

- dans les MNLM, **l'estimation par maximum de vraisemblance** ne peut pas se faire directement car la fonction de vraisemblance n'est en général pas explicite dans ces modèles.
 - dans [11], j'ai travaillé sur l'algorithme MCMC-SAEM et son applicabilité à une classe de modèles de croissance de plantes. En collaboration avec Samis Trevezas (postdoctorant à CentraleSupélec, maintenant Professeur à l'université d'Athènes), nous avons démontré la convergence de l'algorithme pour une famille de modèles de croissance de plantes définis comme des systèmes dynamiques complexes. J'ai également mené l'étude de simulation ainsi que l'analyse des données réelles. J'ai implémenté tous les algorithmes en C++.
 - dans [8], j'ai appliqué la méthodologie précédente à un autre ensemble de données réelles plus riches, en collaboration avec des biologistes. J'ai implémenté le code en C++. Dans [?], j'ai apporté mon soutien à un doctorant (Hussein Kanzo) en statistique appliquée et modélisation, pour l'implémentation de l'algorithme MCMC-SAEM sur un ensemble de données réelles. Ses superviseurs m'ont contactée pour obtenir des conseils sur la procédure d'estimation qu'ils utilisaient, et j'ai suggéré l'utilisation de l'algorithme SAEM. H. Kanzo a implémenté l'algorithme, et nous avons analysé les résultats ensemble.
 - dans [4], en collaboration avec Maud Delattre, Estelle Kuhn, et Jean-Benoist Léger, nous avons proposé un algorithme de descente de gradient stochastique (SGD) pour l'estimation du maximum de vraisemblance dans les modèles à variables latentes, avec des exemples dans les MNLM mais aussi dans les modèles à blocs stochastiques. En utilisant un préconditionneur basé sur la matrice d'information de Fisher, l'algorithme proposé permet une estimation efficace dans de tels modèles, étendant la classe de modèles pour lesquels les algorithmes de type EM sont généralement valides (à savoir, les modèles de la famille exponentielle). J'ai implémenté le code Python pour la partie NLMM.
- dans les MNLM, une question fréquente est de savoir si un effet donné doit être traité comme fixe ou aléatoire. Cette question peut être reformulée en un **test d'hypothèse sur les composantes de la variance**.
 - dans [?], nous avons abordé la question des tests de composantes de la variance dans les MNLM, pour tester si un sous-ensemble des variances est nul contre l'alternative qu'au moins une variance est non nulle. En collaboration avec Paul-Henry Cournède et Estelle Kuhn, nous avons établi que la distribution asymptotique de la statistique du test du rapport de vraisemblance, lorsque la taille de l'échantillon tend vers l'infini, est une loi dite du chi-bar-square (i.e. un mélange de lois du chi-deux), dont les paramètres (les proportions du mélange et les degrés de liberté des composantes du mélange) dépendent en particulier de la matrice d'information de Fisher.

- dans [5], j'ai implémenté les résultats précédents dans un package R. La méthodologie nécessite de savoir simuler selon une loi du chi-bar-square, ainsi qu'une estimation de la matrice d'information de Fisher, par bootstrap paramétrique. Ce package a été téléchargé 27 000 fois et est téléchargé environ 350 fois par mois.
- dans [3], dans le cadre de la thèse de doctorat de Tom Guédon que j'ai co-supervisée avec Estelle Kuhn, nous avons travaillé sur une version à petit échantillon du test du rapport de vraisemblance pour les composantes de la variance. Un test de type bootstrap paramétrique avec seuillage a été proposé, dont les propriétés théoriques et sur simulations ont été étudiées.
- dans [1], nous avons proposé un estimateur de la statistique du test du rapport de vraisemblance basé sur une méthode de type approximation stochastique. Notre méthode peut être utilisée de façon plus générale pour estimer un rapport de constantes de normalisation, mais également pour estimer une unique constante de normalisation. Cela ouvre la voie à des applications en statistique bayésienne, où de tels calculs sont nécessaires, par exemple pour les facteurs de Bayes.

Analyse bayésienne

Pendant mon post-doctorat en collaboration avec des écologues, j'ai travaillé sur l'analyse bayésienne de données complexes.

- dans [9], j'ai proposé un modèle bayésien basé sur des copules pour tenir compte des dépendances entre les observations collectées à plusieurs reprises au cours de deux saisons florales. J'ai implémenté l'algorithme MCMC en R et analysé les résultats.
- dans [6], j'ai travaillé sur un modèle de pollinisation complexe spatialement explicite à l'échelle du paysage, où l'objectif était d'inférer les paramètres du modèle. En raison de la structure hautement non linéaire et de la distribution des données, la vraisemblance n'était pas explicite, ce qui empêchait une analyse bayésienne classique. Ainsi, une approche de type Approximate Bayesian Computation (ABC) a été utilisée.

3.2. Financements, bourses.....

- **Bourse de thèse**, 2025, allocation de la région Hauts-de-France et du CDP C2EMPI de l'Université de Lille, projet co-porté par A. Poulain et O. Bouaziz. **116 000€**.
- **ANR Stat4Plant**, 2021 – 2025, projet porté par E. Kuhn. Je suis responsable du WP1, sur lequel nous avons recruté Tom Guédon. **450 000€**.
- **PEPS JCJC**, 2019 (Projet Exploratoire Premier Soutien, Jeunes Chercheurs, Jeunes Chercheuses) – **3 500€**. Cette bourse m'a permis de financer des déplacements vers la Suède pour finaliser les travaux de recherche que j'avais initiés lors de mon postdoctorat à l'université de Lund.

3.3. Encadrements.....

2025: co-encadrement avec O. Bouaziz (PR ULille) et A. Poulain (CR CNRS ULille) du stage de M2 de Théo Dufresne (Théo poursuit en thèse avec nous à partir de septembre 2025)

2021 - 2024 : co-encadrement avec Estelle Kuhn (DR INRAE) de la thèse de Tom Guédon (recruté MCF au Cérémade (Université PSL))

2021: co-encadrement avec Estelle Kuhn (DR INRAE) du stage de M2 de Tom Guédon

2020: co-encadrement avec Céline Richard-Molard (CR INRAE) du stage de M1 en bio-informatique d'Alexis Koralewski

4. PUBLICATIONS

- [1] T Guédon, C Baey, and E Kuhn. Estimation of ratios of normalizing constants using stochastic approximation: the SARIS algorithm. *Statistics and Computing*, page (to appear), 2025.
- [2] H Kanso, M-M Memah, V Baldazzi, B Quilot-Turion, and C Baey. Assessing inter-individual genetic variability in peach sugar metabolism through reliable parameter estimation of a kinetic model. *European Journal of Agronomy*, page 168, 2025.
- [3] T Guédon, C Baey, and E Kuhn. Bootstrap test procedure for variance components in nonlinear mixed effects models in the presence of nuisance parameters and a singular fisher information matrix. *Biometrika*, page asae025, 2024.
- [4] C Baey, M Delattre, E Kuhn, J-B Leger, and S Lemler. Efficient preconditioned stochastic gradient descent for estimation in latent variable models. In *International Conference on Machine Learning*, pages 1430–1453. PMLR, 2023.
- [5] C Baey and E Kuhn. varTestnlme: An R Package for Variance Components Testing in Linear and Nonlinear Mixed-Effects Models. *Journal of Statistical Software*, 107(6):1–32, 2023.
- [6] C Baey, H G. Smith, M Rundlöf, O Olsson, Y Clough, and U Sahlin. Calibration of a bumble bee foraging model using Approximate Bayesian Computation. *Ecological Modelling*, 477:110251, 2023.
- [7] C Baey, P-H Cournède, and E Kuhn. Asymptotic distribution of likelihood ratio test statistics for variance components in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, 135:107–122, 2019.
- [8] C Baey, A Mathieu, A Jullien, S Trevezas, and P-H Cournède. Modelling inter-individual variability of winter oilseed rape populations using a population-based version of the greenlab model. *Journal of Agricultural, Biological and Environmental Statistics*, 23:208–232, 2018.
- [9] C Baey, U Sahlin, Y Clough, and H G Smith. A model to account for data dependency when estimating floral cover in different land use types over a season. *Environmental and Ecological Statistics*, 24:505–527, 2017.
- [10] J Häussler, U Sahlin, C Baey, H G Smith, and Y Clough. Pollinator population size and pollination ecosystem service responses to enhancing floral and nesting resources. *Ecology and Evolution*, 7:1898–1908, 2017.

- [11] C Baey, S Trevezas, and P-H Cournède. A nonlinear mixed effects model of plant growth and estimation via stochastic variants of the EM algorithm. *Communications in Statistics - Theory and Methods*, 45(6):1643–1669, 2016., 2014.
- [12] C Baey, A Didier, S Lemaire, F Maupas, and P-H Cournède. Modelling the interindividual variability of organogenesis in sugar beet populations using a hierarchical segmented model. *Ecological Modelling*, 263:56–63, 2013.
- [13] C Baey, A Didier, S Lemaire, F Maupas, and P-H Cournède. Parametrization of five classical plant growth models applied to sugar beet and comparison of their predictive capacities on root yield and total biomass. *Ecological Modelling*, 290:11–20, 2013.
- [14] P-H Cournède, Y Chen, Q Wu, C Baey, and B Bayol. Development and Evaluation of Plant Growth Models: Methodology and Implementation in the PyGMAIion platform. *Mathematical Modelling of Natural Phenomena*, 8(4):112–130, 2013.

5. RESPONSABILITÉS ADMINISTRATIVES ET COLLECTIVES

5.1. Animations scientifiques.....

Depuis janvier 2025, je suis chargée de mission égalité du laboratoire. À ce titre, je participe avec deux collègues à la publication d’une newsletter mensuelle sur ce thème, avec des actualités, un “chiffre du mois” mettant en avant une statistique chiffrée en lien avec l’égalité femmes-hommes dans les mathématiques, des suggestions de lectures ou vidéos et un encart historique. Je co-organise également un stage MathCpourL qui aura lieu à Lille en 2026. J’ai également co-organisé deux événements autour de la place des femmes en mathématiques, avec notamment le vernissage de l’exposition photo “Mathématiques, Informatique, avec elles !”. Cette exposition, qui contient 20 portraits de femmes, est prêtée dans les établissements scolaires de la région qui en font la demande. Je coordonne le prêt de l’exposition.

Chaque année, je co-organise avec des collègues du laboratoire une 1/2 journée à destination des étudiant-e-s de licence et master autour des métiers des mathématiques. Le thème change chaque année. À titre d’exemple, j’ai participé aux thématiques “Mathématique et développement durable” (2023), “Mathématique et sécurité” (2024), et la prochaine édition (2025) “Maths et biologie”.

5.2. Responsabilités collectives.....

Depuis 2024, je suis membre élue du CNU section 26.

Depuis 2023, je suis membre du bureau du groupe MAS de la SMAI, et j’en suis également la trésorière.

Depuis 2020, je suis membre élue du conseil de laboratoire du LPP.

J’étais co-organisatrice du séminaire de probabilités et statistiques du laboratoire entre 2018 et 2020.

6. ENSEIGNEMENTS

Le tableau ci-dessous résume les enseignements que j'ai effectués depuis 2017. En 2021–2022, j'ai obtenu une délégation CNRS de 6 mois couplée à mon congé maternité, et n'ai donc pas enseigné.

Année	Intitulé	Type	Cycle	HETD
2024-2025	Statistique Computationnelle	CM+TP	M2 MIASHS	34
	Statistique mathématique	TD	M1 Mathématiques et Applications	30
	Statistique Computationnelle	CM+TD+TP	M1 Mathématiques et Applications	98
	Probabilités et Statistiques	TD	L2 Informatique	18
2023-2024	Statistique mathématique	TD	M1 Mathématiques et Applications	30
	Statistique Computationnelle	CM+TD+TP	M1 Mathématiques et Applications	98
	Algèbre	TD	L2 MIASHS	36
	Probabilités et Statistiques	TD	L2 Informatique	18
2022-2023	Statistique Computationnelle	CM+TD+TP	M1 Mathématiques et Applications	72
	Probabilités et Statistiques	TD	L2 Informatique	18
	Modélisation probabiliste	-	Doctorat	10
2021-2022	<i>décharge pour congé maternité et 1/2 délégation CNRS</i>			-
2020-2021	Méthodes d'apprentissage	CM+TD+TP	M2 ISN	60
	Tests d'hypothèses statistiques	TD	L3 MIASHS	36
	Statistique Computationnelle	CM+TD+TP	M1 Mathématiques et Applications	72
2019-2020	Méthodes d'apprentissage	CM+TD+TP	M2 ISN	72
	Analyse de données	TD+TP	M1 Mathématiques et Applications	30
	Probabilité	TD	L3 Mathématiques	30
	Statistique mathématique	TD	L3 Mathématiques	30
	Tests d'hypothèses statistiques	TD	L3 MIASHS	36
2018-2019	Méthodes d'apprentissage	CM+TD+TP	M2 ISN	72
	Probabilité	TD	L3 Mathématiques	30
	<i>décharge pour congé maternité</i>	-	-	96
2017-2018	Méthodes d'apprentissage	CM+TD+TP	M2 ISN	72
	Analyse de données	TD+TP	M1 Mathématiques et Applications	30
	Probabilité	TD	L3 Mathématiques	30
	Statistique mathématique	TD	L3 Mathématiques	30