

Simon Bussy

CHERCHEUR · ENTREPRENEUR · MACHINE LEARNING PHD

Leading Machine Learning Research @Califrais

06 64 42 71 87 | ✉ simon.bussy@gmail.com | 🏠 www.simonbussy.com | 📱 [SimonBussy](#) | 🌐 [bussysimon](#)

Passionné par l'apprentissage statistique en grande dimension et les mathématiques sous-jacentes, je dirige la recherche au sein de Califrais et suis investi dans différents projets d'enseignement et de recherche théorique et appliquée.

Formations

INSERM

POST-DOCTORAT

Paris, France

Nov. 2018 - Jan. 2020

Introduction de lights : modèle joint d'apprentissage sur données longitudinales en grande dimension dans un contexte d'analyse de survie avec détection de sous-groupes de pronostiques différents.

LPSM, Sorbonne Université

Paris, France

DOCTORAT DE MACHINE LEARNING

Oct. 2015 - Oct. 2018

À l'intersection entre théorie et applications, mon travail a consisté à proposer de nouvelles méthodes interprétables de machine learning en grande dimension, avec un intérêt particulier pour le contexte d'analyse de survie.

ENS Paris-Saclay

Paris, France

M2 RECHERCHE MVA (MENTION TRÈS BIEN)

Sep. 2014 - Aoû. 2015

Harmonic analysis, wavelet analysis and signal processing, optimization, information theory and pattern recognition, statistical learning and high dimensional statistics, compressed sensing, bayesian networks, kernel methods, reinforcement learning, graphical models, computer vision.

Télécom SudParis

Paris, France

INGÉNIEUR, SPÉCIALISÉ EN STATISTIQUE

Sep. 2011 - Aoû. 2014

Maths (A), data analysis (A), probability & statistics (A), data mining (A+), numerical analysis (A), optimization (A), information theory (A+), stochastic processes (A+), Queuing theory (A+), Databases Management (A).

Experiences

Califrais & Foodufrais

Paris, France

CO-FONDATEUR & CTO

Sep. 2014 - Présent

Califrais est une start-up qui révolutionne l'approvisionnement en produits frais et locaux pour les particuliers (foodufrais.fr) et les restaurateurs (califrais.fr) grâce à des outils technologiques avancés, dans un esprit de respect de l'environnement.

- Co-gestion de la stratégie globale de la société (finances, investissements, RH, etc.).
- Co-management de l'équipe (35 personnes).
- Direction du pôle Technique et du Laboratoire de Recherche en Machine Learning (5 personnes).
- Design et développement *from scratch* de la plateforme web : catalogue produits avec mise à jour des données en temps réel, outils de gestion de la logistique et des stocks, outils de prédiction des flux, etc.

INSERM, Centre de Recherche des Cordeliers

Paris, France

CHERCHEUR ASSOCIÉ

Jul. 2019 - Présent

Modélisation statistique et apprentissage, contexte de grande dimension.

LPSM, Sorbonne Université

Paris, France

DOCTORAT DE MACHINE LEARNING

Oct. 2015 - Oct. 2018

Mes travaux de recherche ont principalement porté sur la statistique en grande dimension, avec un intérêt particulier pour le cadre de l'analyse de survie et ses applications pour la médecine personnalisée. J'ai notamment travaillé sur des problèmes liés à des données génétiques ou longitudinales. Mes contributions ont consisté à proposer de nouvelles méthodes dans des études pronostiques pour :

- classer les exemples d'apprentissage dans des sous-groupes relativement à leur risque d'apparition de l'événement d'intérêt,
- détecter automatiquement des seuils dans des covariables continues.

CMAP, École Polytechnique

Palaiseau, France

STAGE DE RECHERCHE EN MACHINE LEARNING

Avr. 2015 - Aoû. 2015

En partant d'une étude de cas sur une cohorte de patients drépanocytaires avec des données longitudinales hétérogènes de grande dimension, l'objectif était de comparer des méthodes d'apprentissage en termes de prédiction du risque et de sélection de variables.

ENS Paris-Saclay

STAGE DE RECHERCHE SUR LES MULTI-ARMED BANDIT

Proposition et étude de nouveaux algorithmes dans un contexte de recommandation de films où le système propose plusieurs choix. Le but du projet était de comprendre ce nouveau contexte, et d'y adapter les algorithmes stochastiques MAB classiques (UCB et Thompson Sampling).

Cachan, France

Nov. 2014 - Jan. 2015

Dreem

PROJET DE RECHERCHE

Le projet, proposé par la startup Dreem, portait sur des données de signaux EEG et le but était de prédire l'apparition d'une oscillation lente. L'idée sous-jacente étant que la stimulation d'une oscillation lente aide au passage d'une phase de sommeil à une autre.

Paris, France

Oct. 2014 - Aoû. 2015

Orange Silicon Valley

STAGE DE DATA SCIENCE

- Test de différentes plateformes Big Data : technologies liées à Hadoop (HBase, HDFS, Pig, Hive, Sqoop), Cassandra et Spark.
- Déploiement de clusters Hadoop sur machines virtuelles et physiques.
- Étude des opportunités commerciales pour les outils Big Data pour les entreprises de télécommunication.
- Développement des modèles de Machine Learning pour prédire l'évolution des tendances de valorisation des start-ups.
- Communication interne autour du *Deep Learning* : fonctionnement & cas d'utilisation.

San Francisco, USA

Fév. 2014 - Aoû. 2014

Télécom SudParis

PROJET DE RECHERCHE EN DEEP LEARNING

Étude théorique & expérimentale de l'article *Layer-Wise Training of Deep Generative Models* (L. Arnold, Y. Ollivier).

Évry, France

Oct. 2013 - Jan. 2014

Télécom SudParis

PROJET DE RECHERCHE

Kalman filter generalization: optimal unsupervised segmentation of pairwise Markov process.

Évry, France

Avr. 2013 - Juin 2013

Thales Communication & Security

STAGE DE RECHERCHE

Performance testing of a seam carving algorithm.

Paris, France

Juin 2012 - Aoû. 2012

Distinctions

2020 **Lauréat**, Prix de thèse Daniel Schwartz 2020, Société Française de Biométrie

Paris, France

2019 **Lauréat**, Prix du Docteur Norbert Marx 2019, Société Française de Statistique

Nancy, France

Papiers de recherche

Lights: a generalized joint model for high-dimensional multivariate longitudinal data and censored durations

S. BUSSY, A. BARBIERI, S. ZOHAR, A.S. JANNOT

2020

Preprint, journal visé : *Journal of the Royal Statistical Society : Series B*

Binacox: automatic cut-points detection in high-dimensional Cox model, with applications to genetic data

S. BUSSY, M.Z. ALAYA, A. GUILLOUX, A.S. JANNOT

2020

Soumis dans *Journal of the American Statistical Association*

[Papier](#), [GitHub](#)

Differential correlation between socio-economic status and clinical characteristics in among men and women in patients consulting in a tertiary hypertension unit.

P. TRAN, A.L. FERAL-PIERSSENS, S. BUSSY, L. AMAR, G. BOBRIE, G. CHATELLIER, A. BURGUN, M. AZIZI, A.S. JANNOT

2020

Preprint, journal visé : *Journal of Hypertension*

Trajectories of biological values and vital parameters: an observational cohort study of non-complicated vaso-occlusive crises

R. VEIL, S. BUSSY, V. LOOTEN, J.B. ARLER, J. POUCHOT CAMOZ, A.S. JANNOT, B. RANQUE

2019

Publié dans *Journal of Clinical Medicine*

[GitHub](#)

Comparison of methods for early-readmission prediction in a high-dimensional heterogeneous covariates and time-to-event outcome framework

S. BUSSY, R. VEIL, V. LOOTEN, A. BURGUN, S. GAÏFFAS, A. GUILLOUX, B. RANQUE, A.S. JANNOT

2019

Publié dans *BMC Medical Research Methodology*

[Papier](#), [GitHub](#)

Binarsity: a penalization for one-hot encoded features in linear supervised learning

M.Z. ALAYA, S. BUSSY, S. GAÏFFAS, A. GUILLOUX

2019

Publié dans *Journal of Machine Learning Research*

[Papier](#), [GitHub](#)

C-mix: A high-dimensional mixture model for censored durations, with applications to genetic data

S. BUSSY, A. GUILLOUX, S. GAÏFFAS, A.S. JANNOT

Publié dans *Statistical Methods in Medical Research*

[Papier](#), [GitHub](#)

2019

Enseignements

2019-21	Apprentissage et Données Longitudinales , M2 Données Massives en Santé - Université Paris Descartes	Paris, France
2019-21	Survival Analysis et Machine Learning , M2 Données Massives en Santé - Université Paris Descartes	Paris, France
2018-21	DataCamp Kaggle , Encadrement des projets, M2 Data Science - Université Paris Saclay	Évry, France
2019	Data Science & Feature Design , Ecole polytechnique Executive Education	Palaiseau, France
2016-18	Modèles linéaires , TD, M1 ISUP-UPMC	Paris, France
2015-18	Statistique inférentielle , TD, L3 ISUP-UPMC	Paris, France
2015-16	Séries temporelles , TD/TP, M1 ISUP-UPMC	Paris, France

Exposés

Introduction de modèles de machine learning interprétables en grande dimension et leurs applications	Paris, France
PRIX DE THÈSE DANIEL SCHWARTZ - CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS	Octobre 2020
Lights: a generalized joint model for high-dimensional multivariate longitudinal data and censored durations	Philadelphia, USA
2020 JOINT STATISTICAL MEETINGS (JSM) - AMERICAN STATISTICAL ASSOCIATION (oral, virtual session)	Août 2020
Lights: a generalized joint model for high-dimensional multivariate longitudinal data and censored durations	Seoul, Corée
2020 INTERNATIONAL BIOMETRIC CONFERENCE (IBC) (oral, virtual session)	Juillet 2020
Prix du Docteur Norbert MARX 2019	Nancy, France
JOURNÉES DE STATISTIQUE 2019 (oral)	Juin 2019
How to introduce AI to a traditional business	Paris, France
IRONTALK: CONFÉRENCE À IRONHACK WEB DEVELOPMENT BOOTCAMP (oral)	Mai. 2019
Grand témoin docteur-entrepreneur: Business Plan, Valeur d'une entreprise et Financement de l'innovation	Paris, France
CONFÉRENCE DOCTORIALES SORBONNE UNIVERSITÉ (oral)	Avr. 2019
Design of a machine learning algorithm in high dimension to predict hospital readmission	Paris, France
SÉMINAIRE INSERM UMR 1137 (PARIS BICHAT) (oral)	Fév. 2019
Design d'un algorithme d'IA en grande dimension pour prédire la réadmission à l'hôpital	Paris, France
SÉMINAIRE IRDES MÉTHODOLOGIQUE SUR LES PARCOURS DE SOINS (oral)	Jan. 2019
Binacox: automatic cut-points detection in high-dimensional Cox model	Évry, France
L'INTELLIGENCE ARTIFICIELLE POUR L'HOMME ET SA SANTÉ (poster)	Déc. 2018
Design d'un algorithme d'IA en grande dimension pour prédire la réadmission à l'hôpital	Nancy, France
I.A. & SANTÉ (oral)	Juil. 2018
Binacox: automatic cut-points detection in high-dimensional Cox model	Buenos Aires, Argentine
MACHINE LEARNING SUMMER SCHOOL (poster)	Juin 2018
Survival model in high-dimension	Paris, France
SÉMINAIRE DES DOCTORANTS DU LPSM (oral)	Avr. 2017
C-mix: A high-dimensional mixture model for censored durations	Birmingham, United Kingdom
INTERNATIONAL SOCIETY FOR CLINICAL BIostatISTICS (poster)	Août. 2016
Modèle pénalisé pronostique à variable latente pour des données censurées	Paris, France
SÉMINAIRE DU MAP5 (poster)	Mai 2016

A high dimensional mixture model for time-to-event data

PARIS BIG DATA MANAGEMENT SUMMIT (poster)

Paris, France

Mar. 2016

Modelling Patient Time-Series Data from EHR using Gaussian Processes

WORKSHOP DATA INITIATIVE, ÉCOLE POLYTECHNIQUE (oral)

Palaiseau, France

Mar. 2016

Facteurs prédictifs de la réhospitalisation précoce de drépanocytaires adultes

SÉMINAIRE DU DÉPARTEMENT DE MÉDECINE INTERNE DE L'HEGP (oral)

Paris, France

Jan. 2016

Risk assessment of sickle-cell anemia

SÉMINAIRE DES DOCTORANTS DU LPSM (oral)

Paris, France

Jan. 2016

Continuous time survival in latent variable models

SÉMINAIRE DES DOCTORANTS DU CENTRE DE RECHERCHE DES CORDELIERS (oral)

Paris, France

Déc. 2015

Forecasting Non-Stationary Time Series

WORKSHOP DATA INITIATIVE, ÉCOLE POLYTECHNIQUE (oral)

Palaiseau, France

Oct. 2015

Compétences techniques

Back-end Python, PHP, REST API, SQL, MySQL

Front-end HTML, CSS, Javascript

Programmation Python, R, C++, JAVA, LaTeX

Langues Français, Anglais, Espagnol

Encadrement

Van Tuan Nguyen - Califrais Machine Learning Lab

STAGE DE RECHERCHE, M2 TÉLÉCOM PARIS

High-dimensional constrained stock optimization using deep deterministic policy gradient

Paris, France

2020

Alice Eliet-Doillet - Califrais Machine Learning Lab

STAGE DE RECHERCHE, M2 ENSAE PARISTECH

Machine Learning for churn modelling: state-of-the-art review and introduction of a new method for high-dimensional longitudinal data

Paris, France

2019

Bastien Bergere - Califrais Machine Learning Lab

STAGE DE RECHERCHE, M2 MATHÉMATIQUES, APPRENTISSAGE ET SCIENCES HUMAINES (MASH) - UNIVERSITÉ

PARIS-DAUPHINE; MATHÉMATIQUES VISION APPRENTISSAGE (MVA) - ENS PARIS-SACLAY

Reinforcement Learning for stock optimization and introduction of a new method for learning an efficient discretization for continuous states MDP

Paris, France

2019

Meziane Bellahmer - Califrais Machine Learning Lab

STAGE DE RECHERCHE, M2 INTELLIGENCE ARTIFICIELLE - UNIVERSITÉ PARIS-DESCARTES

Multivariate time-series modelling to predict prices, availability and demand of fresh produce

Paris, France

2019

Maud De Tollenaere - co-encadrée avec A.S. Jannot, F. Pages et A. Guilloux

STAGE DE RECHERCHE, M2 ANALYSE, MODÉLISATION ET INGÉNIERIE DE L'INFORMATION BIOLOGIQUE ET MÉDICALE -

UNIVERSITÉ PARIS-SACLAY

Techniques d'apprentissage automatique pour affiner le pronostic à l'aide de données d'imagerie immunologique

Paris, France

2018

Christophe Botella - co-encadré avec A.S. Jannot et A. Guilloux

STAGE DE RECHERCHE, M2 MATHÉMATIQUES POUR LES SCIENCES DU VIVANT - UNIVERSITÉ PARIS-SACLAY

Modèle de régression sur données longitudinales: application à des données médicales à faible résolution

Paris, France

2016