Activity Report 2018

Project-Team MISTIS

Modelling and Inference of Complex and Structured Stochastic Systems

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

Table of contents

Project-Team MISTIS

Creation of the Project-Team: 2008 January 01

Keywords:

Computer Science and Digital Science:

A3.1.1. - Modeling, representation
A3.1.4. - Uncertain data
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A3.4.5. - Bayesian methods
A3.4.7. - Kernel methods
A5.3.3. - Pattern recognition
A5.9.2. - Estimation, modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A6.3.3. - Data processing
A6.3.5. - Uncertainty Quantification
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

B1.2.1. - Understanding and simulation of the brain and the nervous system
B2.6.1. - Brain imaging
B3.3. - Geosciences
B3.4.1. - Natural risks
B3.4.2. - Industrial risks and waste
B3.5. - Agronomy
B5.1. - Factory of the future
B9.5.6. - Data science
B9.11.1. - Environmental risks

1. Team, Visitors, External Collaborators

Research Scientists

Julyan Arbel [Inria, Researcher]
Florence Forbes [Team leader, Inria, Senior Researcher, HDR]
Stéphane Girard [Inria, Senior Researcher, HDR]

Faculty Member

Jean-Baptiste Durand [Institut polytechnique de Grenoble, Associate Professor]

Post-Doctoral Fellows

Jean-Michel Bécu [Inria, until Mar 2018]

Marta Crispino [Inria, from Mar 2018]

Pascal Dkengne Sielenou [Inria, from Nov 2018]

Hongliang Lu [Inria]

Antoine Usseglio Carleve [Inria, from Oct 2018]

Fei Zheng [Inria, from Mar 2018]

PhD Students

Clément Albert [Inria]

Alexis Arnaud [Institut polytechnique de Grenoble]

Karina Ashurbekova [Univ Grenoble Alpes]

Meryem Bousebata [Univ Grenoble Alpes, from Oct 2018]

Fabien Boux [Univ Grenoble Alpes]

Alexandre Constantin [Univ Grenoble Alpes, from Oct 2018]

Benoit Kugler [Univ Grenoble Alpes, from Sep 2018]

Veronica Munoz Ramirez [Univ Grenoble Alpes]

Brice Olivier [Univ Grenoble Alpes]

Thibaud Rahier [Cifre Schneider Electric, until Nov 2018]

Pierre-Antoine Rodesch [CEA, until Oct 2018]

Mariia Vladimirova [Inria, from Oct 2018]

Technical staff

Jaime Eduardo Arias Almeida [Inria, until Jan 2018]

Fatima Fofana [Inria]

Steven Quinito Masnada [Inria, until Aug 2018]

Interns

Virgilio Kmetzsch Rosa E Silva [Inria, until Jun 2018]

Benoit Kugler [Univ Grenoble Alpes, from Apr 2018 until Aug 2018]

Caroline Lawless [Inria, from Jun 2018 until Jul 2018]

Michal Lewandowski [Inria, from Feb 2018 until Jul 2018]

Yassine Meftah [Inria, from May 2018 until Jul 2018]

Theo Moins [Inria, from May 2018 until Aug 2018]

Mariia Vladimirova [Inria, from Feb 2018 until Jul 2018]

Administrative Assistant

Marion Ponsot [Inria]

Visiting Scientists

Aboubacrène Ag Ahmad [Univ. Gaston Berger, Senegal, from Sep 2018 until Nov 2018]

Hien Nguyen [Inria, from Sep 2018 until Oct 2018]

External Collaborator

Eric Paul Marchand [Univ. Sherbrook, Canada, from Mar 2018 until Jun 2018]

2. Overall Objectives

2.1. Overall Objectives

The context of our work is the analysis of structured stochastic models with statistical tools. The idea underlying the concept of structure is that stochastic systems that exhibit great complexity can be accounted for by combining simple local assumptions in a coherent way. This provides a key to modelling, computation, inference and interpretation. This approach appears to be useful in a number of high impact applications including signal and image processing, neuroscience, genomics, sensors networks, etc. while the needs from these domains can in turn generate interesting theoretical developments. However, this powerful and flexible approach can still be restricted by necessary simplifying assumptions and several generic sources of complexity in data.

Often data exhibit complex dependence structures, having to do for example with repeated measurements on individual items, or natural grouping of individual observations due to the method of sampling, spatial or temporal association, family relationship, and so on. Other sources of complexity are related to the measurement process, such as having multiple measuring instruments or simulations generating high dimensional and heterogeneous data or such that data are dropped out or missing. Such complications in data-generating processes raise a number of challenges. Our goal is to contribute to statistical modelling by offering theoretical concepts and computational tools to handle properly some of these issues that are frequent in modern data. So doing, we aim at developing innovative techniques for high scientific, societal, economic impact applications and in particular via image processing and spatial data analysis in environment, biology and medicine.

The methods we focus on involve mixture models, Markov models, and more generally hidden structure models identified by stochastic algorithms on one hand, and semi and non-parametric methods on the other hand.

Hidden structure models are useful for taking into account heterogeneity in data. They concern many areas of statistics (finite mixture analysis, hidden Markov models, graphical models, random effect models, ...). Due to their missing data structure, they induce specific difficulties for both estimating the model parameters and assessing performance. The team focuses on research regarding both aspects. We design specific algorithms for estimating the parameters of missing structure models and we propose and study specific criteria for choosing the most relevant missing structure models in several contexts.

Semi and non-parametric methods are relevant and useful when no appropriate parametric model exists for the data under study either because of data complexity, or because information is missing. When observations are curves, they enable us to model the data without a discretization step. These techniques are also of great use for *dimension reduction* purposes. They enable dimension reduction of the functional or multivariate data with no assumptions on the observations distribution. Semi-parametric methods refer to methods that include both parametric and non-parametric aspects. Examples include the Sliced Inverse Regression (SIR) method which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis*, which is based on the modelling of distribution tails by both a functional part and a real parameter.

3. Research Program

3.1. Mixture models

Participants: Alexis Arnaud, Jean-Baptiste Durand, Florence Forbes, Stéphane Girard, Julyan Arbel, Jean-Michel Bécu, Hongliang Lu, Fabien Boux, Veronica Munoz Ramirez, Benoit Kugler, Alexandre Constantin, Fei Zheng.

Key-words: mixture of distributions, EM algorithm, missing data, conditional independence, statistical pattern recognition, clustering, unsupervised and partially supervised learning.

In a first approach, we consider statistical parametric models, θ being the parameter, possibly multi-dimensional, usually unknown and to be estimated. We consider cases where the data naturally divides into observed data $y = \{y_1, ..., y_n\}$ and unobserved or missing data $z = \{z_1, ..., z_n\}$. The missing data z_i represents for instance the memberships of one of a set of K alternative categories. The distribution of an observed y_i can be written as a finite mixture of distributions,

$$f(y_i; \theta) = \sum_{k=1}^{K} P(z_i = k; \theta) f(y_i \mid z_i; \theta) . \tag{1}$$

These models are interesting in that they may point out hidden variables responsible for most of the observed variability and so that the observed variables are *conditionally* independent. Their estimation is often difficult due to the missing data. The Expectation-Maximization (EM) algorithm is a general and now standard approach to maximization of the likelihood in missing data problems. It provides parameter estimation but also values for missing data.

Mixture models correspond to independent z_i's. They have been increasingly used in statistical pattern recognition. They enable a formal (model-based) approach to (unsupervised) clustering.

3.2. Markov models

Participants: Alexis Arnaud, Brice Olivier, Thibaud Rahier, Jean-Baptiste Durand, Florence Forbes, Karina Ashurbekova, Hongliang Lu, Pierre-Antoine Rodesch, Julyan Arbel, Mariia Vladimirova.

Key-words: graphical models, Markov properties, hidden Markov models, clustering, missing data, mixture of distributions, EM algorithm, image analysis, Bayesian inference.

Graphical modelling provides a diagrammatic representation of the dependency structure of a joint probability distribution, in the form of a network or graph depicting the local relations among variables. The graph can have directed or undirected links or edges between the nodes, which represent the individual variables. Associated with the graph are various Markov properties that specify how the graph encodes conditional independence assumptions.

It is the conditional independence assumptions that give graphical models their fundamental modular structure, enabling computation of globally interesting quantities from local specifications. In this way graphical models form an essential basis for our methodologies based on structures.

The graphs can be either directed, e.g. Bayesian Networks, or undirected, e.g. Markov Random Fields. The specificity of Markovian models is that the dependencies between the nodes are limited to the nearest neighbor nodes. The neighborhood definition can vary and be adapted to the problem of interest. When parts of the variables (nodes) are not observed or missing, we refer to these models as Hidden Markov Models (HMM). Hidden Markov chains or hidden Markov fields correspond to cases where the z_i's in (1) are distributed according to a Markov chain or a Markov field. They are a natural extension of mixture models. They are widely used in signal processing (speech recognition, genome sequence analysis) and in image processing (remote sensing, MRI, etc.). Such models are very flexible in practice and can naturally account for the phenomena to be studied.

Hidden Markov models are very useful in modelling spatial dependencies but these dependencies and the possible existence of hidden variables are also responsible for a typically large amount of computation. It follows that the statistical analysis may not be straightforward. Typical issues are related to the neighborhood structure to be chosen when not dictated by the context and the possible high dimensionality of the observations. This also requires a good understanding of the role of each parameter and methods to tune them depending on the goal in mind. Regarding estimation algorithms, they correspond to an energy minimization problem which is NP-hard and usually performed through approximation. We focus on a certain type of methods based on variational approximations and propose effective algorithms which show good performance in practice and for which we also study theoretical properties. We also propose some tools for model selection. Eventually we investigate ways to extend the standard Hidden Markov Field model to increase its modelling power.

3.3. Functional Inference, semi- and non-parametric methods

Participants: Clément Albert, Stéphane Girard, Florence Forbes, Jean-Michel Bécu, Antoine Usseglio Carleve, Pascal Dkengne Sielenou, Marta Crispino, Meryem Bousebata.

Key-words: dimension reduction, extreme value analysis, functional estimation.

We also consider methods which do not assume a parametric model. The approaches are non-parametric in the sense that they do not require the assumption of a prior model on the unknown quantities. This property is important since, for image applications for instance, it is very difficult to introduce sufficiently general parametric models because of the wide variety of image contents. Projection methods are then a way to decompose the unknown quantity on a set of functions (*e.g.* wavelets). Kernel methods which rely on smoothing the data using a set of kernels (usually probability distributions) are other examples. Relationships exist between these methods and learning techniques using Support Vector Machine (SVM) as this appears in the context of *level-sets estimation* (see section 3.3.2). Such non-parametric methods have become the cornerstone when dealing with functional data [77]. This is the case, for instance, when observations are curves. They enable us to model the data without a discretization step. More generally, these techniques are of great use for *dimension reduction* purposes (section 3.3.3). They enable reduction of the dimension of the functional or multivariate data without assumptions on the observations distribution. Semi-parametric methods refer to methods that include both parametric and non-parametric aspects. Examples include the Sliced Inverse Regression (SIR) method [79] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [76], which is based on the modelling of distribution tails (see section 3.3.1). It differs from traditional statistics which focuses on the central part of distributions, *i.e.* on the most probable events. Extreme value theory shows that distribution tails can be modelled by both a functional part and a real parameter, the extreme value index.

3.3.1. *Modelling extremal events*

Extreme value theory is a branch of statistics dealing with the extreme deviations from the bulk of probability distributions. More specifically, it focuses on the limiting distributions for the minimum or the maximum of a large collection of random observations from the same arbitrary distribution. Let $X_{1,n} \leq ... \leq X_{n,n}$ denote n ordered observations from a random variable X representing some quantity of interest. A p_n-quantile of X is the value x_{p_n} such that the probability that X is greater than x_{p_n} is p_n, *i.e.* $P(X > x_{p_n}) = p_n$. When $p_n < 1/n$, such a quantile is said to be extreme since it is usually greater than the maximum observation $X_{n,n}$ (see Figure 1).

To estimate such quantiles therefore requires dedicated methods to extrapolate information beyond the observed values of X. Those methods are based on Extreme value theory. This kind of issue appeared in hydrology. One objective was to assess risk for highly unusual events, such as 100-year floods, starting from flows measured over 50 years. To this end, semi-parametric models of the tail are considered:

$$P(X > x) = x^{-1/\theta}\ell(x), \ x > x_0 > 0, \tag{2}$$

where both the extreme-value index $\theta > 0$ and the function $\ell(x)$ are unknown. The function ℓ is a slowly varying function *i.e.* such that

$$\frac{\ell(tx)}{\ell(x)} \to 1 \ \text{as} \ x \to \infty \tag{3}$$

for all $t > 0$. The function $\ell(x)$ acts as a nuisance parameter which yields a bias in the classical extreme-value estimators developed so far. Such models are often referred to as heavy-tail models since the probability of extreme events decreases at a polynomial rate to zero. It may be necessary to refine the model (2,3) by specifying a precise rate of convergence in (3). To this end, a second order condition is introduced involving an additional parameter $\rho \leq 0$. The larger ρ is, the slower the convergence in (3) and the more difficult the estimation of extreme quantiles.
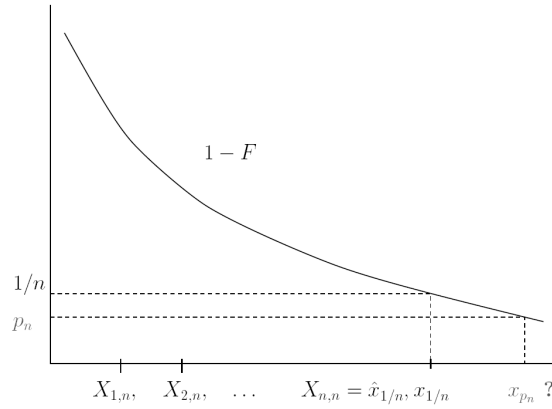
Figure 1. The curve represents the survival function $x \to P(X > x)$. The $1/n$-quantile is estimated by the maximum observation so that $\widehat{x}_{1/n} = X_{n,n}$. As illustrated in the figure, to estimate p_n-quantiles with $p_n < 1/n$, it is necessary to extrapolate beyond the maximum observation.

More generally, the problems that we address are part of the risk management theory. For instance, in reliability, the distributions of interest are included in a semi-parametric family whose tails are decreasing exponentially fast. These so-called Weibull-tail distributions [9] are defined by their survival distribution function:

$$P(X > x) = \exp\left\{-x^{\theta}\ell(x)\right\},\ x > x_0 > 0. \tag{4}$$

Gaussian, gamma, exponential and Weibull distributions, among others, are included in this family. An important part of our work consists in establishing links between models (2) and (4) in order to propose new estimation methods. We also consider the case where the observations were recorded with a covariate information. In this case, the extreme-value index and the p_n-quantile are functions of the covariate. We propose estimators of these functions by using moving window approaches, nearest neighbor methods, or kernel estimators.

3.3.2. Level sets estimation

Level sets estimation is a recurrent problem in statistics which is linked to outlier detection. In biology, one is interested in estimating reference curves, that is to say curves which bound 90% (for example) of the population. Points outside this bound are considered as outliers compared to the reference population. Level sets estimation can be looked at as a conditional quantile estimation problem which benefits from a non-parametric statistical framework. In particular, boundary estimation, arising in image segmentation as well as in supervised learning, is interpreted as an extreme level set estimation problem. Level sets estimation can also be formulated as a linear programming problem. In this context, estimates are sparse since they involve only a small fraction of the dataset, called the set of support vectors.

3.3.3. Dimension reduction

Our work on high dimensional data requires that we face the curse of dimensionality phenomenon. Indeed, the modelling of high dimensional data requires complex models and thus the estimation of high number of parameters compared to the sample size. In this framework, dimension reduction methods aim at replacing the original variables by a small number of linear combinations with as small as a possible loss of information.

Principal Component Analysis (PCA) is the most widely used method to reduce dimension in data. However, standard linear PCA can be quite inefficient on image data where even simple image distortions can lead to highly non-linear data. Two directions are investigated. First, non-linear PCAs can be proposed, leading to semi-parametric dimension reduction methods [78]. Another field of investigation is to take into account the application goal in the dimension reduction step. One of our approaches is therefore to develop new Gaussian models of high dimensional data for parametric inference [75]. Such models can then be used in a Mixtures or Markov framework for classification purposes. Another approach consists in combining dimension reduction, regularization techniques, and regression techniques to improve the Sliced Inverse Regression method [79].

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Veronica Munoz Ramirez, Florence Forbes, Stéphane Girard, Jaime Eduardo Arias Almeida, Pierre-Antoine Rodesch, Hongliang Lu, Fabien Boux, Benoit Kugler, Alexandre Constantin.

As regards applications, several areas of image analysis can be covered using the tools developed in the team. More specifically, in collaboration with team PERCEPTION, we address various issues in computer vision involving Bayesian modelling and probabilistic clustering techniques. Other applications in medical imaging are natural. We work more specifically on MRI and functional MRI data, in collaboration with the Grenoble Institute of Neuroscience (GIN) and the NeuroSpin center of CEA Saclay. We also consider other statistical 2D fields coming from other domains such as remote sensing, in collaboration with the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) and the Centre National d'Etudes Spatiales (CNES). In this context, we worked on hyperspectral and/or multitemporal images. In the context of the "pole de competivité" project I-VP, we worked of images of PC Boards. We also address reconstruction problems in tomography with CEA Grenoble.

4.2. Multi sensor Data Analysis

Participants: Jean-Michel Bécu, Florence Forbes, Thibaud Rahier, Hongliang Lu, Fatima Fofana.

A number of our methods are at the the intersection of data fusion, statistics, machine learning and acoustic signal processing. The context can be the surveillance and monitoring of a zone acoustic state from data acquired at a continuous rate by a set of sensors that are potentially mobile and of different nature (eg WIFUZ project with the ACOEM company in the context of a DGA-rapid initiative). Typical objectives include the development of prototypes for surveillance and monitoring that are able to combine multi sensor data coming from acoustic sensors (microphones and antennas) and optical sensors (infrared cameras) and to distribute the processing to multiple algorithmic blocs. Our interest in acoustic data analysis mainly started from past European projects, POP and Humavips, in collaboration with the PERCEPTION team (PhD theses of Vassil Khalidov, Ramya Narasimha, Antoine Deleforge, Xavier Alameda, and Israel Gebru).

4.3. Biology, Environment and Medicine

Participants: Jaime Eduardo Arias Almeida, Alexis Arnaud, Florence Forbes, Stéphane Girard, Jean-Baptiste Durand, Clément Albert, Julyan Arbel, Jean-Michel Bécu, Thibaud Rahier, Brice Olivier, Karina Ashurbekova, Fabien Boux, Veronica Munoz Ramirez, Fei Zheng.

A third domain of applications concerns biology and medicine. We considered the use of mixture models to identify biomakers. We also investigated statistical tools for the analysis of fluorescence signals in molecular biology. Applications in neurosciences are also considered. In the environmental domain, we considered the modelling of high-impact weather events and the use of hyperspectral data as a new tool for quantitative ecology.

5. Highlights of the Year

5.1. Highlights of the Year

Scholarships:

- Alexandre Constantin supervised by S. Girard (MISTIS) and M. Fauvel (INRA Toulouse) was granted a PhD scholarship on "Analyse de séries temporelles massives d'images satellitaires: Applications à la cartographie des écosystèmes" from CNES and the IDEX Université Grenoble Alpes – Initiatives de Recherche Stratégiques (IRS).

- Meryem Bousebata supervised by S. Girard (MISTIS) and G. Enjolras (CERAG Grenoble) was granted a PhD scholarship on "Bayesian estimation of extreme risk measures: Implication for the insurance of natural disasters" from the Idex project named Risk@UGA.

Projects:

- In the context of another Idex project named Data@UGA, a 2-year multi-disciplinary project entitled "Tracking and analysis of large population of dynamic single molecules" was granted in November 2018 to MISTIS in collaboration with the GIN, coordinated by F. Forbes (MISTIS) and V. Stoppin-Mellet (GIN).

Editorial and publishing activities:

- A new book entitled *Handbook of mixture analysis*, edited at CRC Press by Gilles Celeux (Inria), Sylvia Früwirth-Schnatter (Wien University), and Christian P. Robert (Université Paris-Dauphine) is now available (December 2018). Florence Forbes and Julyan Arbel have written 2 of the chapters in the book [49], [51].

- Marianne Clausel and Jean-Baptiste Durand co-published a chapter [48] on generative models in data science in the book *Data Science. Cours et exercices*, edited by Eyrolles (Paris).

- Stéphane Girard and Julyan Arbel have co-edited a book of proceedings following the Summer School Stat4Astro they organized in Autrans in 2017 [64].

New appointments:

- Stéphane Girard has been hired as a research collaborator by the CMAP (Centre de Mathématiques Appliquées de l'école Polytechnique) in the context of the Chair Stress Test, RISK Management and Financial Steering, led by the French Ecole polytechnique and its Foundation and sponsored by BNP Paribas.

6. New Software and Platforms

6.1. BOLD model FIT

KEYWORDS: Functional imaging - FMRI - Health

SCIENTIFIC DESCRIPTION: Physiological and biophysical models have been proposed to link neuronal activity to the Blood Oxygen Level-Dependent (BOLD) signal in functional MRI (fMRI). Those models rely on a set of parameter values that are commonly estimated using gradient-based local search methods whose initial values are taken from the literature. In some applications, interesting insight into the brain physiology or physiopathology can be gained from an estimation of the model parameters from measured BOLD signals. In this work we focus on the extended Balloon model and propose the estimation of 15 parameters using seven different approaches: three versions of the Expectation Maximization Gauss-Newton (EM/GN) approach (the *de facto* standard in the neuroscientific community) and four metaheuristics (Particle Swarm Optimization (PSO), Differential Evolution (DE), Real-Coded Genetic Algorithms (GA), and a Memetic Algorithm (MA) combining EM/GN and DE). To combine both the ability to escape local optima and to incorporate prior knowledge, we derive the target function from Bayesian modeling. The general behavior of these algorithms

is analyzed and compared, providing very promising results on challenging real and synthetic fMRI data sets involving rats with epileptic activity. These stochastic optimizers provided a better performance than EM/GN in terms of distance to the ground truth in 4 out of 6 synthetic data sets and a better signal fitting in 12 out of 12 real data sets. Non-parametric statistical tests showed the existence of statistically significant differences between the real data results obtained by DE and EM/GN. Finally, the estimates obtained from DE for these parameters seem both more realistic and more stable or at least as stable across sessions as the estimates from EM/GN. This is the largest comparison of optimizers for the estimation of biophysical parameters in BOLD fMRI

FUNCTIONAL DESCRIPTION: This Matlab toolbox performs the automatic estimation of biophysical parameters using the extended Balloon model and BOLD fMRI data. It takes as input a MAT file and provides as output the parameter estimates achieved by using stochastic optimization

NEWS OF THE YEAR: The main differences with our previous work: 1) we also use synthetic data, 2) we use stochastic GN and MCMC+DE, 3) We evaluate results not only in physiological terms but also comparing fitness function values. Also changes were made to allow running on the cluster via MPI

- Participants: Pablo Mesejo Santiago, Florence Forbes and Jan Warnking
- Partner: University of Granada, Spain
- Contact: Pablo Mesejo Santiago
- Publication: A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data
- URL: https://hal.archives-ouvertes.fr/hal-01221115v2/

6.2. PyHRF

KEYWORDS: Medical imaging - Health - Brain - IRM - Neurosciences - Statistic analysis - FMRI

SCIENTIFIC DESCRIPTION: Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that allows the non-invasive study of brain function. It is based on the hemodynamic variations induced by changes in cerebral synaptic activity following sensory or cognitive stimulation. The measured signal depends on the variation of blood oxygenation level (BOLD signal) which is related to brain activity: a decrease in deoxyhemoglobin concentration induces an increase in BOLD signal. The BOLD signal is delayed with respect to changes in synaptic activity, which can be modeled as a convolution with the Hemodynamic Response Function (HRF) whose exact form is unknown and fluctuates with various parameters such as age, brain region or physiological conditions. In this work we propose to analyze fMRI data using a Joint Detection-Estimation (JDE) approach. It jointly detects cortical activation and estimates the HRF. In contrast to existing tools, PyHRF estimates the HRF instead of considering it as a given constant in the entire brain.

FUNCTIONAL DESCRIPTION: As part of fMRI data analysis, PyHRF provides a set of tools for addressing the two main issues involved in intra-subject fMRI data analysis : (i) the localization of cerebral regions that elicit evoked activity and (ii) the estimation of the activation dynamics also referenced to as the recovery of the Hemodynamic Response Function (HRF). To tackle these two problems, PyHRF implements the Joint Detection-Estimation framework (JDE) which recovers parcel-level HRFs and embeds an adaptive spatio-temporal regularization scheme of activation maps.

NEWS OF THE YEAR: The framework to perform software tests has been further developed. Some unitary tests have been set.

- Participants: Aina Frau Pascual, Christine Bakhous, Florence Forbes, Jaime Eduardo Arias Almeida, Laurent Risser, Lotfi Chaari, Philippe Ciuciu, Solveig Badillo, Thomas Perret and Thomas Vincent
- Partners: CEA - NeuroSpin
- Contact: Florence Forbes
- Publications: Frontiers in Neuroinformatics Flexible multivariate hemodynamics fMRI data analyses and simulations with PyHRF - Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach - A Bayesian Non-Parametric Hidden Markov Random Model for Hemodynamic Brain Parcellation
- URL: http://pyhrf.org

6.3. xLLiM

High dimensional locally linear mapping
KEYWORDS: Clustering - Regression
SCIENTIFIC DESCRIPTION: Building a regression model for the purpose of prediction is widely used in all disciplines. A large number of applications consists of learning the association between responses and predictors and focusing on predicting responses for the newly observed samples. In this work, we go beyond simple linear models and focus on predicting low-dimensional responses using high-dimensional covariates when the associations between responses and covariates are non-linear.
FUNCTIONAL DESCRIPTION: This is an R package available on the CRAN at https://cran.r-project.org/web/packages/xLLiM/index.html

XLLiM provides a tool for non linear mapping (non linear regression) using a mixture of regression model and an inverse regression strategy. The methods include the GLLiM model (Deleforge et al (2015)) based on Gaussian mixtures and a robust version of GLLiM, named SLLiM (see Perthame et al (2016)) based on a mixture of Generalized Student distributions.
NEWS OF THE YEAR: A new Hierarchical version of GLLiM has been developed in collaboration with University of Michigan, USA.

- Participants: Antoine Deleforge, Emeline Perthame and Florence Forbes

- Partner: University of Michigan, Ann Arbor, USA

- Contact: Florence Forbes

- Publications: Inverse regression approach to robust nonlinear high-to-low dimensional mapping - High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables

- URL: https://cran.r-project.org/web/packages/xLLiM/index.html

6.4. MMST

Mixtures of Multiple Scaled Student T distributions
KEYWORDS: Health - Statistics - Brain MRI - Medical imaging - Robust clustering
SCIENTIFIC DESCRIPTION: A new family of multivariate heavy-tailed distributions that allow variable marginal amounts of tailweight is proposed and implemented. The originality comes from introducing multidimensional instead of univariate scale variables for the mixture of scaled Gaussian family of distributions. In contrast to most existing approaches, the derived distributions can account for a variety of shapes and have a simple tractable form with a closed-form probability density function whatever the dimension. We provide maximum likelihood estimation of the parameters and illustrate their modelling flexibility.
FUNCTIONAL DESCRIPTION: The package implements mixtures of so-called multiple scaled Student distributions, which are generalisation of multivariate Student T distribution allowing different tails in each dimension. Typical applications include Robust clustering to analyse data with possible outliers. In this context, the model and package have been used on large data sets of brain MRI to segment and identify brain tumors. Recent additions include a Markov random field implementation to account for spatial dependencies between observations, and a Bayesian implementation that can be used to select the number of mixture components automatically.

RELEASE FUNCTIONAL DESCRIPTION: Recent additions include a Markov random field implementation to account for spatial dependencies between observations, and a Bayesian implementation that can be used to select the number of mixture components automatically.

NEWS OF THE YEAR: Recent additions include a Markov random field implementation to account for spatial dependencies between observations, and a Bayesian implementation that can be used to select the number of mixture components automatically.

- Participants: Alexis Arnaud, Darren Wraith, Florence Forbes, Steven Quinito Masnada and Stéphane Despréaux

- Partner: Institut des Neurosciences Grenoble

- Contact: Florence Forbes

- Publications: A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering - Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric Quantitative MRI Data

- URL: https://team.inria.fr/mistis/software/

7. New Results

7.1. Mixture models

7.1.1. *Hierarchical mixture of linear mappings in high dimension*

Participant: Florence Forbes.

Joint work with: Benjamin Lemasson from Grenoble Institute of Neuroscience, Naisyin Wang and Chun-Chen Tu from University of Michigan, Ann Arbor, USA.

Regression is a widely used statistical tool. A large number of applications consists of learning the association between responses and predictors. From such an association, different tasks, including prediction, can then be conducted. To go beyond simple linear models while maintaining tractability, non-linear mappings can be handled through exploration of local linearity. The non-linear relationship can be captured by a mixture of locally linear regression models as proposed in the so-called Gaussian Locally Linear Mapping (GLLiM) model [6] that assumes Gaussian noise models. In the past year, we have been working on several extensions and applications of GLLiM as described below and the next two subsections.

We proposed a structured mixture model called Hierarchical Locally Linear Mapping (HGLLiM), to predict low-dimensional responses based on high dimensional covariates when the associations between the responses and the covariates are non-linear. For tractability, HGLLiM adopts inverse regression to handle the high dimension and locally-linear mappings to capture potentially non-linear relations. Data with similar associations are grouped together to form a cluster. A mixture is composed of several clusters following a hierarchical structure. This structure enables shared covariance matrices and latent factors across smaller clusters to limit the number of parameters to estimate. Moreover, HGLLiM adopts a robust estimation procedure for model stability. We used three real-world datasets to demonstrate different features of HGLLiM. With the face dataset, HGLLiM shows the ability of modeling non-linear relationship through mixtures. With the orange juice dataset, we show the prediction performance of HGLLiM is robust to the presence of outliers. Moreover, we demonstrated that HGLLiM is capable of handling large-scale complex data using the data acquired from a magnetic resonance vascular fingerprinting (MRvF) study. These examples illustrate the wide applicability of HGLLiM on handling different aspects of a complex data structure in prediction. A preliminary version of this work under revision for JRSS-C can be found in [72].

7.1.2. *Dictionary-free MR fingerprinting parameter estimation via inverse regression*

Participants: Florence Forbes, Fabien Boux, Julyan Arbel.

Joint work with: Emmanuel Barbier from Grenoble Institute of Neuroscience.

Magnetic resonance imaging (MRI) can map a wide range of tissue properties but is often limited to observe a single parameter at a time. In order to overcome this problem, Ma et al. introduced magnetic resonance fingerprinting (MRF), a procedure based on a dictionary of simulated couples of signals and parameters. Acquired signals called fingerprints are then matched to the closest signal in the dictionary in order to estimate parameters. This requires an exhaustive search in the dictionary, which even for moderately sized problems, becomes costly and possibly intractable . We propose an alternative approach to estimate more parameters at a time. Instead of an exhaustive search for every signal, we use the dictionary to learn the functional relationship between signals and parameters. This allows the direct estimation of parameters without the need of searching through the dictionary. We investigated the use of GLLiM [6] that bypasses the problems associated with high-to-low regression. The experimental validation of our method is performed in the context of vascular fingerprinting. The comparison between a standard grid search and the proposed approach suggest that MR Fingerprinting could benefit from a regression approach to limit dictionary size and fasten computation time. Preliminary tests and results have been presented at International Society for Magnetic Resonance in Medicine conference, ISMRM 2018 [35].

7.1.3. *Massive analysis of multi-angular hyperspectral images of the planet Mars by inverse regression of physical models*

Participants: Florence Forbes, Benoit Kugler.

Joint work with: Sylvain Douté from Institut de Planétologie et d'Astrophysique de Grenoble (IPAG).

In the starting PhD of Benoit Kugler, the objective is to develop a statistical learning technique capable of solving a complex inverse problem in planetary remote sensing. The challenges are 1) the large number of observations to to inverse, 2) their large dimension, 3) the need to provide predictions for correlated parameters and 4) the need to provide a quality index (eg. uncertainty). To achieve this goal, we have started to investigate a setting in which a physical model is available to provide simulations that can then be used for learning prior to inversion of real observed data. For the learning step to be as accurate as possible, an initial task is then to estimate the best fit of the theoretical model to the real data. We proposed an iterative procedure based on a combination of GLLiM [6] predictions and importance sampling steps.

7.1.4. *Quantitative MRI Characterization of Brain Abnormalities in de novo Parkinsonian patients*

Participants: Florence Forbes, Veronica Munoz Ramirez, Alexis Arnaud, Julyan Arbel.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience.

Currently there is an important delay between the onset of Parkinson's disease and its diagnosis. The detection of changes in physical properties of brain structures may help to detect the disease earlier. In this work, we proposed to take advantage of the informative features provided by quantitative MRI to construct statistical models representing healthy brain tissues. We used mixture models of non Gaussian distributions [8] to capture the non-standard shape of the data multivariate distribution. This allowed us to detect atypical values for these features in the brain of Parkinsonian patients following a procedure similar to that in [16]. Promising preliminary results demonstrate the potential of our approach in discriminating patients from controls and revealing the subcortical structures the most impacted by the disease. This work has been accepted at the IEEE International Symposium on Biological Imaging, ISBI 2019 [36].

7.1.5. *No structural differences are revealed by voxel-based morphometry in de novo Parkinsonian patients*

Participants: Florence Forbes, Veronica Munoz Ramirez.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience and Pierrick Coupé from Laboratoire Bordelais de Recherche en Informatique, UMR 5800, Univ. Bordeaux, Talence.

The identification of brain morphological alterations in newly diagnosed PD patients (i.e. de novo) could potentially serve as a biomarker and accelerate diagnosis. However, presently no consensus exists in the literature possibly due to several factors: small size cohorts, differences in segmentation techniques or bad control of false positive rates. In this study, we seek, using the Computational Anatomy Toolbox (CAT12) (University of Jena) pipeline, for morphological brain differences in gray and white matter of 66 controls and 144 de novo PD patients whose data were extracted from the PPMI (Parkinson Progressive Markers Initiative) database. Moreover, we searched for subcortical structure differences using the new online platform VolBrain (J. V. Manjón and P. Coupé, "volBrain: An Online MRI Brain Volumetry System," Front. Neuroinform., vol. 10, p. 30, Jul. 2016). We found no structural brain differences in this de novo Parkinsonian population, neither in tissues using a whole brain analysis nor in any of nine subcortical structures analyzed separately. We concluded that some results published in the literature appear as false positives and are not reproducible.

7.1.6. *Characterization of daily glycemic variability in the patient with type 1 diabetes*

Participants: Florence Forbes, Fei Zheng.

Joint work with: Stéphane Bonnet from CEA Leti and Pierre-Yves Benhamou, Manon Jalbert from CHU Grenoble Alpes.

Glycemic variability (GV) is an important component of glycemic control in patients with type 1 diabetes. Many metrics have been proposed to account for this variability but none is unanimous among physicians. One difficulty is that the variations in blood sugar levels are expressed very differently from one day to another in some subjects. Our goal was to develop and evaluate the performance of a daily GV index built by combining different known metrics (CV, MAGE, GVP etc). This in order to merge their descriptive power to obtain a more complete and more accurate index. This preliminary study will be presented at the Société Francophone du Diabète (SFD) in 2019 [46].

7.1.7. *Glycemic variability improves after pancreatic islet transplantation in patients with type 1 diabetes*

Participants: Florence Forbes, Fei Zheng.

Joint work with: Stéphane Bonnet from CEA Leti and Pierre-Yves Benhamou, Manon Jalbert from CHU Grenoble Alpes.

Glycemic variability (GV) must be taken into account in the efficacy of treatment of type 1 diabetes because it determines the quality of glycemic control, the risk of complication of the patient's disease. Our goal in this study was to describe GV scores in patients with pancreatic islet transplantation (PIT) type 1 diabetes in the TRIMECO trial, and change of thresholds, for each index. predictive of success of PIT.

7.1.8. *Dirichlet process mixtures under affine transformations of the data*

Participant: Julyan Arbel.

Joint work with: Riccardo Corradin from Milano Bicocca, Italy and Bernardo Nipoti from Trinity College Dublin, Ireland.

Location-scale Dirichlet process mixtures of Gaussians (DPM-G) have proved extremely useful in dealing with density estimation and clustering problems in a wide range of domains. Motivated by an astronomical application, in this work we address the robustness of DPM-G models to affine transformations of the data, a natural requirement for any sensible statistical method for density estimation. In [57], we first devise a coherent prior specification of the model which makes posterior inference invariant with respect to affine transformation of the data. Second, we formalize the notion of asymptotic robustness under data transformation and show that mild assumptions on the true data generating process are sufficient to ensure that DPM-G models feature such a property. As a by-product, we derive weaker assumptions than those provided in the literature for ensuring posterior consistency of Dirichlet process mixtures, which could reveal of independent interest. Our investigation is supported by an extensive simulation study and illustrated by the analysis of an astronomical dataset consisting of physical measurements of stars in the field of the globular cluster NGC 2419.

7.1.9. Applications of mixture models in Industry

Participant: Julyan Arbel.

Joint work with: Kerrie Mengersen, Earl Duncan, Clair Alston-Knox and Nicole White.

A very wide range of commonly encountered problems in industry are amenable to statistical mixture modelling and analysis. These include process monitoring or quality control, efficient resource allocation, risk assessment, prediction, and so on. Commonly articulated reasons for adopting a mixture approach include the ability to describe non-standard outcomes and processes, the potential to characterize each of a set of multiple outcomes or processes via the mixture components, the concomitant improvement in interpretability of the results, and the opportunity to make probabilistic inferences such as component membership and overall prediction.

In [51], We illustrate the wide diversity of applications of mixture models to problems in industry, and the potential advantages of these approaches, through a series of case studies.

7.1.10. Approximation results regarding the multiple-output mixture of the Gaussian-gated linear experts model

Participant: Florence Forbes.

Joint work with: Hien Nguyen, La Trobe University Melbourne Australia and Faicel Chamroukhi, Caen University, France.

Mixture of experts (MoE) models are a class of artificial neural networks that can be used for functional approximation and probabilistic modeling. An important class of MoE models is the class of mixture of linear experts (MoLE) models, where the expert functions map to real topological output spaces. Recently, Gaussian-gated MoLE models have become popular in applied research. There are a number of powerful approximation results regarding Gaussian-gated MoLE models, when the output space is univariate. These results guarantee the ability of Gaussian-gated MoLE mean functions to approximate arbitrary continuous functions, and Gaussian-gated MoLE models themselves to approximate arbitrary conditional probability density functions. We utilized and extended upon the univariate approximation results in order to prove a pair of useful results for situations where the output spaces are multivariate. We do this by proving a pair of lemmas regarding the combination of univariate MoLE models, which are interesting in their own rights.

7.1.11. Models for ranking data

Participant: Marta Crispino.

within the BigInsight project, Oslo.

We developed a new method and algorithms for working with ranking data. This kind of data is particularly relevant in applications involving personalized recommendations. In particular, we have invented a new Bayesian approach based on extensions of the Mallows model, which allows making personalized recommendations equipped with a level of uncertainty.

The Mallows model (MM) is a popular parametric family of models for ranking data, based on the assumption that a modal ranking, which can be interpreted as the consensus ranking of the population, exists. The probability of observing a given ranking is then assumed to decay exponentially fast as its distance from the consensus grows. The MM is therefore a two-parameter distance-based family of models. The scale or precision parameter, controlling the concentration of the distribution determines the rate of decay of the probability of individual ranks. Individual models with different properties can be obtained depending on the choice of distance on the space of permutations. A major drawback of the MM is that its computational complexity has limited its use to a particular form based on Kendall distance. We develop new computationally tractable methods for Bayesian inference in Mallows models that work with any right-invariant distance. Our method performs inference on the consensus ranking of the items, also when based on partial rankings, such as top-k items or pairwise comparisons. When assessors are many or heterogeneous, we propose a mixture model for clustering them in homogeneous subgroups, with cluster specific consensus rankings. We develop

approximate stochastic algorithms that allow a fully probabilistic analysis, leading to coherent quantifications of uncertainties, make probabilistic predictions on the class membership of assessors based on their ranking of just some items, and predict missing individual preferences, as needed in recommendation systems. The methodology has been published in the Journal of Machine Learning Research, JMLR, in early 2018.

A generalization of the model above involves dealing with non-transitive and heterogeneous pairwise comparison data, coming from an experiment within the musicology domain. We thus develop a mixture model extension of the Bayesian Mallows model able to handle non-transitive data, with a latent layer of uncertainty which captures the generation of preference misreporting. This paper was recently accepted for publication in the Annals of Applied Statistics, AoAS.

Within this project, we also write a survey paper, whose main goal is to compare the performance of our method with other existing methodologies, including the Plackett-Luce, the Bradley-Terry, the collaborative filtering methods, and some of their variations. We illustrate and discuss the use of these models by means of an experiment in which assessors rank potatoes, and with a simulation. The purpose of this paper is not to recommend the use of one best method, but to present a palette of different possibilities for different questions and different types of data. This was recently accepted on the Annual Review of Statistics and Its Applications, ARSIA.

7.2. Semi and non-parametric methods

7.2.1. *Estimation of extreme risk measures*

Participant: Stéphane Girard.

Joint work with: A. Daouia (Univ. Toulouse), L. Gardes (Univ. Strasbourg), J. Elmethni (Univ. Paris 5) and G. Stupfler (Univ. Nottingham, UK).

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0, 1)$ corresponds to the upper α-quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: $\mathrm{VaR}(\alpha)$ does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α-quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution. In [20], we investigate the extreme properties of a new risk measure (called the Conditional Tail Moment) which encompasses various risk measures, such as the CTE, as particular cases. We study the situation where some covariate information is available under some general conditions on the distribution tail. We thus has to deal with conditional extremes. However, the asymptotic normality of the empirical CTE estimator requires that the underlying distribution possess a finite variance; this can be a strong restriction in heavy-tailed models which constitute the favoured class of models in actuarial and financial applications. One possible solution in very heavy-tailed models where this assumption fails could be to use the more robust Median Shortfall, but this quantity is actually just a quantile, which therefore only gives information about the frequency of a tail event and not about its typical magnitude. In [65], we construct a synthetic class of tail L_p -medians, which encompasses the Median Shortfall (for $p = 1$) and Conditional Tail Expectation (for $p = 2$). We show that, for $1 < p < 2$, a tail L_p -median always takes into account both the frequency and magnitude of tail events, and its empirical estimator is, within the range of the data, asymptotically normal under a condition weaker than a finite variance. We extrapolate this estimator, along with another technique, to proper extreme levels using the heavy-tailed framework. The estimators are showcased on a simulation study and on a set of real fire insurance data showing evidence of a very heavy right tail.

A possible coherent alternative risk measure is based on expectiles [18], [63], [62]. Compared to quantiles, the family of expectiles is based on squared rather than absolute error loss minimization. The flexibility and virtues of these least squares analogues of quantiles are now well established in actuarial science, econometrics and statistical finance. Both quantiles and expectiles were embedded in the more general class of M-quantiles [19] as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles.

7.2.2. *Extrapolation limits associated with extreme-value methods*

Participants: Clément Albert, Stéphane Girard.
Joint work with: L. Gardes (Univ. Strasbourg) and A. Dutfoy (EDF R&D).

The PhD thesis of Clément Albert (co-funded by EDF) is dedicated to the study of the sensitivity of extreme-value methods to small changes in the data and to their extrapolation ability. Two directions are explored:

(i) In [54], we investigate the asymptotic behavior of the (relative) extrapolation error associated with some estimators of extreme quantiles based on extreme-value theory. It is shown that the extrapolation error can be interpreted as the remainder of a first order Taylor expansion. Necessary and sufficient conditions are then provided such that this error tends to zero as the sample size increases. Interestingly, in case of the so-called Exponential Tail estimator, these conditions lead to a subdivision of Gumbel maximum domain of attraction into three subsets. In contrast, the extrapolation error associated with Weissman estimator has a common behavior over the whole Fréchet maximum domain of attraction. First order equivalents of the extrapolation error are then derived and their accuracy is illustrated numerically.

(ii) In [53], We propose a new estimator for extreme quantiles under the log-generalized Weibull-tail model, introduced by Cees de Valk. This model relies on a new regular variation condition which, in some situations, permits to extrapolate further into the tails than the classical assumption in extreme-value theory. The asymptotic normality of the estimator is established and its finite sample properties are illustrated both on simulated and real datasets.

7.2.3. *Estimation of local intrinsic dimensionality with extreme-value methods*

Participant: Stéphane Girard.
Joint work with: L. Amsaleg (LinkMedia, Inria Rennes), O. Chelly (NII Japon), T. Furon (LinkMedia, Inria Rennes), M. Houle (NII Japon), K.-I. Kawarabayashi (NII Japon), M. Nett (Google).

This work is concerned with the estimation of a local measure of intrinsic dimensionality (ID). The local model can be regarded as an extension of Karger and Ruhl's expansion dimension to a statistical setting in which the distribution of distances to a query point is modeled in terms of a continuous random variable. This form of intrinsic dimensionality can be particularly useful in search, classification, outlier detection, and other contexts in machine learning, databases, and data mining, as it has been shown to be equivalent to a measure of the discriminative power of similarity functions. In [14], several estimators of local ID are proposed and analyzed based on extreme value theory, using maximum likelihood estimation, the method of moments, probability weighted moments, and regularly varying functions. An experimental evaluation is also provided, using both real and artificial data.

7.2.4. *Bayesian inference for copulas*

Participants: Julyan Arbel, Marta Crispino, Stéphane Girard.

We study in [58] a broad class of asymmetric copulas known as Liebscher copulas and defined as a combination of multiple—usually symmetric—copulas. The main thrust of this work is to provide new theoretical properties including exact tail dependence expressions and stability properties. A subclass of Liebscher copulas obtained by combining Fréchet copulas is studied in more details. We establish further dependence properties for copulas of this class and show that they are characterized by an arbitrary number of singular components. Furthermore, we introduce a novel iterative construction for general Liebscher copulas which *de facto* insures uniform margins, thus relaxing a constraint of Liebscher's original construction. Besides, we show that this iterative construction proves useful for inference by developing an Approximate Bayesian computation sampling scheme. This inferential procedure is demonstrated on simulated data.

In [22], we investigate the properties of a new transformation of copulas based on the co-copula and an univariate function. It is shown that several families in the copula literature can be interpreted as particular outputs of this transformation. Symmetry, association, ordering and dependence properties of the resulting copula are established.

7.2.5. *Bayesian nonparametric clustering*

Participant: Julyan Arbel.

Joint work with: Riccardo Corradin from Milano Bicocca, Michal Lewandowski from Bocconi University, Milan, Italy, Caroline Lawless from Université Paris-Dauphine, France.

For a long time, the Dirichlet process has been the gold standard discrete random measure in Bayesian nonparametrics. The Pitman–Yor process provides a simple and mathematically tractable generalization, allowing for a very flexible control of the clustering behaviour. Two commonly used representations of the Pitman–Yor process are the stick-breaking process and the Chinese restaurant process. The former is a constructive representation of the process which turns out very handy for practical implementation, while the latter describes the partition distribution induced. Obtaining one from the other is usually done indirectly with use of measure theory. In contrast, we propose in [66] an elementary proof of Pitman–Yor's Chinese Restaurant process from its stick-breaking representation.

In the discussion paper [56], we propose a simulation study to emphasise the difference between Variation of Information and Binder's loss functions in terms of number of clusters estimated by means of the use of the Markov chain Monte Carlo output only and a "greedy" method.

The chapter [47] is part of a book edited by Stéphane Girard and Julyan Arbel. It presents a Bayesian nonparametric approach to clustering, which is particularly relevant when the number of components in the clustering is unknown. The approach is illustrated with the Milky Way's globulars, that are clouds of stars orbiting in our galaxy. Clustering globulars is key for better understanding the Milky Way's history. We define the Dirichlet process and illustrate some alternative definitions such as the Chinese restaurant process, the Pólya Urn, the Ewens sampling formula, the stick-breaking representation through some simple *R* code. The Dirichlet process mixture model is presented, as well as the *R* package *BNPmix* implementing Markov chain Monte Carlo sampling. Inference for the clustering is done with the variation of information loss function.

7.2.6. *Multi sensor fusion for acoustic surveillance and monitoring*

Participants: Florence Forbes, Jean-Michel Bécu.

Joint work with: Pascal Vouagner and Christophe Thirard from ACOEM company.

In the context of the DGA-rapid WIFUZ project, we addressed the issue of determining the localization of shots from multiple measurements coming from multiple sensors. The WIFUZ project is a collaborative work between various partners: DGA, ACOEM and HIKOB companies and Inria. This project is at the intersection of data fusion, statistics, machine learning and acoustic signal processing. The general context is the surveillance and monitoring of a zone acoustic state from data acquired at a continuous rate by a set of sensors that are potentially mobile and of different nature. The overall objective is to develop a prototype for surveillance and monitoring that is able to combine multi sensor data coming from acoustic sensors (microphones and antennas) and optical sensors (infrared cameras) and to distribute the processing to multiple algorithmic blocs. As an illustration, the MISTIS contribution is to develop technical and scientific solutions as part of a collaborative protection approach, ideally used to guide the best coordinated response between the different vehicles of a military convoy. Indeed, in the case of an attack on a convoy, identifying the threatened vehicles and the origin of the threat is necessary to organize the best response from all members on the convoy. Thus it will be possible to react to the first contact (emergency detection) to provide the best answer for threatened vehicles (escape, lure) and for those not threatened (suppression fire, riposte fire). We developed statistical tools that make it possible to analyze this information (characterization of the threat) using fusion of acoustic and image data from a set of sensors located on various vehicles. We used Bayesian inversion and simulation techniques to recover multiple sources mimicking collaborative interaction between several vehicles.

7.2.7. *Extraction and data analysis toward "industry of the future"*
Participants: Florence Forbes, Hongliang Lu, Fatima Fofana, Jaime Eduardo Arias Almeida.

Joint work with: J. F. Cuccaro and J. C Trochet from Vi-Technology company.

Industry as we know it today will soon disappear. In the future, the machines which constitute the manufacturing process will communicate automatically as to optimize its performance as whole. Transmitted information essentially will be of statistical nature. In the context of VISION 4.0 project with Vi-Technology, the role of MISTIS is to identify what statistical methods might be useful for the printed circuits boards assembly industry. The topic of F. Fofana's internship was to extract and analyze data from two inspection machines of a industrial process making electronic cards. After a first extraction step in the SQL database, the goal was to enlighten the statistical links between these machines. Preliminary experiments and results on the Solder Paste Inspection (SPI) step, at the beginning of the line, helped identifying potentially relevant variables and measurements (eg related to stencil offsets) to identify future defects and discriminate between them. More generally, we have access to two databases at both ends (SPI and Component Inspection) of the assembly process. The goal is to improve our understanding of interactions in the assembly process, find out correlations between defects and physical measures, generate proactive alarms so as to detect departures from normality.

7.2.8. *Change point detection for the analysis of dynamic single molecules*
Participants: Florence Forbes, Theo Moins.

Joint work with: Virginie Stoppin-Mellet from Grenoble Institute of Neuroscience.

The objective of this study was to develop a statistical learning technique to analyze signals produced by molecules. The main difficulties are the noisy nature of the signals and the definition of a quality index to allow the elimination of poor-quality data and false positive signals. In collaboration with the GIN, we addressed the statistical analysis of intensity traces (2 month internship of Theo Moins, Ensimag 2A). Namely, the ImageJ Thunderstorm toolbox, which has been developed for the detection of single molecule in super resolution imaging, has been successfully used to detect immobile single molecules and generate time-dependent intensity traces. Then the R package Segmentor3IsBack, a fast segmentation algorithm based on 5 possible statistical models, proved efficient in the processing of the noisy intensity traces. This preliminary study led to a multidisciplinary project funded by the Grenoble data institute for 2 years in which we will also address additional challenges for the tracking of a large population of single molecules.

7.3. Graphical and Markov models

7.3.1. *Fast Bayesian network structure learning using quasi-determinism screening*
Participants: Thibaud Rahier, Stéphane Girard, Florence Forbes.

Joint work with: Sylvain Marié, Schneider Electric.

Learning the structure of Bayesian networks from data is a NP-Hard problem that involves an optimization task on a super-exponential sized space. In this work, we show that in most real life datasets, a number of the arcs contained in the final structure can be prescreened at low computational cost with a limited impact on the global graph score. We formalize the identification of these arcs via the notion of quasi-determinism, and propose an associated algorithm that reduces the structure learning to a subset of the original variables. We show, on diverse benchmark datasets, that this algorithm exhibits a significant decrease in computational time and complexity for only a little decrease in performance score. A first version of this work can be found in [71] and has been presented at the JFRB 2018 workshop [41].

7.3.2. *Robust structure learning using multivariate t-distributions*
Participants: Karina Ashurbekova, Florence Forbes.

Joint work with: Sophie Achard, senior researcher at CNRS, Gipsa-lab.

Structure learning is an active topic nowadays in different application areas, i.e. genetics, neuroscience. We addressed the issue of robust graph structure learning in continuous settings. We focused on sparse precision matrix estimation for its tractability and ability to reveal some measure of dependence between variables. For this purpose, we proposed to extract good features from existing methods, namely *tlasso* and CLIME procedures. The former is based on the observation that standard Gaussian modelling results in procedures that are too sensitive to outliers and proposes the use of t-distributions as an alternative. The latter is an alternative to the popular Lasso optimization principle which can handle some of its limitations. We then combined these ideas into a new procedure referred to as tCLIME that can be seen as a modified *tlasso* algorithm. Numerical performance was investigated using simulated data and reveals that tCLIME performs favorably compared to the other standard methods. This work was presented at the Journées de Statistiques de la Société Francaise de Statistique in Saclay, 2018, [39].

7.3.3. *Structure learning via Hadamard product of correlation and partial correlation matrices*
Participants: Karina Ashurbekova, Florence Forbes.

Joint work with: Sophie Achard, senior researcher at CNRS, Gipsa-lab.

Classical conditional independences or marginal independences may not be sufficient to express complex relationships. In this work we introduced a new structure learning procedure where an edge in the graph corresponds to a non zero of both correlation and partial correlation. A theoretical study was derived which shows the good properties of the proposed graph estimator, illustrated also on a synthetic example.

7.3.4. *Spatial mixtures of multiple scaled t-distributions*
Participants: Florence Forbes, Alexis Arnaud.

Joint work with: Steven Quinito Masnada, Inria Grenoble Rhone-Alpes

The goal is to implement an hidden Markov model version of our recently introduced mixtures of non standard multiple scaled t-distributions. The motivation for doing that is the application to multiparametric MRI data for lesion analysis. When dealing with MRI human data, spatial information is of primary importance. For our preliminary study on rat data [16], the results without spatial information were already quite smooth. The main anatomical structures can be identified. We suspect the reason is that the measured parameters already contain a lot of information about the underlying tissues. However, introducing spatial information is always useful and is our ongoing work. In the statistical framework we have developed (mixture models and EM algorithm), it is conceptually straightforward to introduce an additional Markov random field. In addition, when using a Markov random field it is easy to incorporate additional atlas information.

7.3.5. *Spectral CT reconstruction with an explicit photon-counting detector model: a "one-step" approach*
Participants: Florence Forbes, Pierre-Antoine Rodesch.

Joint work with: Veronique Rebuffel and Clarisse Fournier from CEA-LETI Grenoble.

In the context of Pierre-Antoine Rodesh's PhD thesis, we investigate new statistical and optimization methods for tomographic reconstruction from non standard detectors providing multiple energy signals. Recent developments in energy-discriminating Photon-Counting Detector (PCD) enable new horizons for spectral CT. With PCDs, new reconstruction methods take advantage of the spectral information measured through energy measurement bins. However PCDs have serious spectral distortion issues due to charge-sharing, fluorescence escape, pileup effect Spectral CT with PCDs can be decomposed into two problems: a noisy geometric inversion problem (as in standard CT) and an additional PCD spectral degradation problem. The aim of this study is to introduce a reconstruction method which solves both problems simultaneously: a one-step approach. An explicit linear detector model is used and characterized by a Detector Response Matrix (DRM). The algorithm reconstructs two basis material maps from energy-window transmission data. The results prove that the simultaneous inversion of both problems is well performed for simulation data. For comparison, we also perform a standard two-step approach: an advanced polynomial decomposition of measured sinograms combined with a filtered-back projection reconstruction. The results demonstrate the potential uses of this method for medical imaging or for non-destructive control in industry. Preliminary results have been presented at the SPIE medical imaging 2018 conference in Houston, USA [37].

7.3.6. *Non parametric Bayesian priors for hidden Markov random fields*

Participants: Florence Forbes, Julyan Arbel, Hongliang Lu.

Hidden Markov random field (HMRF) models are widely used for image segmentation or more generally for clustering data under spatial constraints. They can be seen as spatial extensions of independent mixture models. As for standard mixtures, one concern is the automatic selection of the proper number of components in the mixture, or equivalently the number of states in the hidden Markov field. A number of criteria exist to select this number automatically based on penalized likelihood (eg. AIC, BIC, ICL etc.) but they usually require to run several models for different number of classes to choose the best one. Other techniques (eg. reversible jump) use a fully Bayesian setting including a prior on the class number but at the cost of prohibitive computational times. In this work, we investigate alternatives based on the more recent field of Bayesian nonparametrics. In particular, Dirichlet process mixture models (DPMM) have emerged as promising candidates for clustering applications where the number of clusters is unknown. Most applications of DPMM involve observations which are supposed to be independent. For more complex tasks such as unsupervised image segmentation with spatial relationships or dependencies between the observations, DPMM are not satisfying. This work has been presented at the Joint Statistical Meeting in Vancouver Canada [29] and at the Journées de la Statistique in Saclay [40].

7.3.7. *Hidden Markov models for the analysis of eye movements*

Participants: Jean-Baptiste Durand, Brice Olivier.

This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.
Joint work with: Anne Guérin-Dugué (GIPSA-lab) and Benoit Lemaire (Laboratoire de Psychologie et Neurocognition)

In the last years, GIPSA-lab has developed computational models of information search in web-like materials, using data from both eye-tracking and electroencephalograms (EEGs). These data were obtained from experiments, in which subjects had to decide whether a text was related or not to a target topic presented to them beforehand. In such tasks, reading process and decision making are closely related. Statistical analysis of such data aims at deciphering underlying dependency structures in these processes. Hidden Markov models (HMMs) have been used on eye movement series to infer phases in the reading process that can be interpreted as steps in the cognitive processes leading to decision. In HMMs, each phase is associated with a state of the Markov chain. The states are observed indirectly though eye-movements. Our approach was inspired by Simola et al. (2008), but we used hidden semi-Markov models for better characterization of phase length distributions [80]. The estimated HMM highlighted contrasted reading strategies (ie, state transitions), with both individual and document-related variability. However, the characteristics of eye movements within each phase tended to be poorly discriminated. As a result, high uncertainty in the phase changes arose, and it could be difficult to relate phases to known patterns in EEGs.

This is why, as part of Brice Olivier's PhD thesis, we are developed integrated models coupling EEG and eye movements within one single HMM for better identification of the phases. Here, the coupling incorporates some delay between the transitions in both (EEG and eye-movement) chains, since EEG patterns associated to cognitive processes occur lately with respect to eye-movement phases. Moreover, EEGs and scanpaths were recorded with different time resolutions, so that some resampling scheme had to be added into the model, for the sake of synchronizing both processes. An associated EM algorithm for maximum likelihood parameter estimation was derived.

New results were obtained in the standalone analysis of the eye-movements. A comparison between the effects of three types of texts was performed, considering texts either closely related, moderately related or unrelated to the target topic.

Our goal for this coming year is to implement and validate our coupled model for jointly analyzing eye-movements and EEGs in order to improve the discrimination of the reading strategies.

7.3.8. *Lossy compression of tree structures*

Participant: Jean-Baptiste Durand.

Joint work with: Christophe Godin and Romain Azaïs (Inria Mosaic)

The class of self-nested trees presents remarkable compression properties because of the systematic repetition of subtrees in their structure. The aim of our work is to achieve compression of any unordered tree by finding the nearest self-nested tree. Solving this optmization problem without more assumptions is conjectured to be an NP-complete or NP-hard problem. In [34], we firstly provided a better combinatorial characterization of this specific family of trees. In particular, we showed from both theoretical and practical viewpoints that complex queries can be quickly answered in self-nested trees compared to general trees. We also presented an approximation algorithm of a tree by a self-nested one that can be used in fast prediction of edit distance between two trees.

Our goal for this coming year is to apply this approach to quantify the degree of self-nestedness of several plant species and extend first results obtained on rice panicles stating that near self-nestedness is a fairly general pattern in plants.

7.3.9. *Relations between structural characteristics in rose bush and visual sensory attributes for objective evaluation of the visual quality*

Participant: Jean-Baptiste Durand.

Joint work with: Gilles Galopin (QUASAV, Agrocampus Ouest)

Within ornamental horticulture context, visual quality of plants is a critical criterion for consumers looking for immediate decorative effect products. Studying links between architecture and its phenotypic plasticity in response to growing conditions and the resulting plant visual appearance represents an interesting lever to propose a new approach for managing product quality from specialized crops. Objectives of the present study were to determine whether architectural components may be identified across different growing conditions (1) to study the architectural development of a shrub over time; and (2) to predict sensory attributes data characterizing multiple visual traits of the plants. The approach addressed in this study stands on the sensory profile method using a recurrent blooming modern rose bush presented in rotation using video stimuli. Plants were cultivated under a shading gradient in three distinct environments (natural conditions, under 55% and 75% shading nets). Architecture and video of the plants were recorded during three stages, from 5 to 15 months after plant multiplication. Predictive models of visual quality were obtained with regression and variable transformation to encompass non-linear relationships [21]. The proposed approach is a way to gain a better insight into the architecture of shrub plants together with their visual appearance to target processes of interest in order to optimize growing conditions or select the most fitting genotypes across breeding programs, with respect to contrasted consumer preferences.

As a perspective, dynamic traits issued from hidden-Markov-based growth models should be used for a better characterization of visual quality, as well as identification of reiterated complexes, which are believed to play a major role in rose bush structure.

7.3.10. Bayesian neural networks

Participants: Julyan Arbel, Mariia Vladimirova.

Joint work with: Pablo Mesejo from University of Granada, Spain.

We investigate in [45] and [44] deep Bayesian neural networks with Gaussian priors on the weights and ReLU-like nonlinearities, shedding light on novel sparsity-inducing mechanisms at the level of the units of the network, both pre- and post-nonlinearities. The main thrust of the paper is to establish that the units prior distribution becomes increasingly heavy-tailed with depth. We show that first layer units are Gaussian, second layer units are sub-Exponential, and we introduce sub-Weibull distributions to characterize the deeper layers units. Bayesian neural networks with Gaussian priors are well known to induce the weight decay penalty on the weights. In contrast, our result indicates a more elaborate regularisation scheme at the level of the units. This result provides new theoretical insight on deep Bayesian neural networks, underpinning their natural shrinkage properties and practical potential.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

CIFRE PhD with SCHNEIDER (2015-2018).   F. Forbes and S. Girard are the advisors of a CIFRE PhD (T. Rahier) with Schneider Electric. The other advisor is S. Marié from Schneider Electric. The goal is to develop specific data mining techniques able to merge and to take advantage of both structured and unstructured (meta)data collected by a wide variety of Schneider Electric sensors to improve the quality of insights that can be produced. The total financial support for MISTIS is of 165 keuros.

PhD contract with EDF (2016-2018).   S. Girard is the advisor of a PhD (A. Clément) with EDF. The goal is to investigate sensitivity analysis and extrapolation limits in extreme-value theory with application to extreme weather events. The financial support for MISTIS is of 140 keuros.

Contract with VALEO.   S. Girard and Pascal Dkengne Sielenou are involved in a study with Valeo to assess the relevance of extreme-value theory in the calibration of sensors for autonomous cars. The financial support for MISTIS is of 100 keuros.

Contract with Andritz.   F. Forbes and C. Braillon (SED) are involved in a study with Andritz to elaborate metrics based on image analysis to assess the quality of nonwaven tissues. The financial support for MISTIS is of 15 keuros.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Grenoble Idex projects

MISTIS is involved in a transdiciplinary project **NeuroCoG** and in a newly accepted cross-disciplinary project (CDP) **Risk@UGA**. F. Forbes is also a member of the executive committee and responsible for the *Data Science for life sciences* work package in another project entitled **Grenoble Alpes Data Institute**.

- The main objective of the RISK@UGA project is to provide some innovative tools both for the management of risk and crises in areas that are made vulnerable because of strong interdependencies between human, natural or technological hazards, in synergy with the conclusions of Sendai conference. The project federates a hundred researchers from Human and Social Sciences, Information & System Sciences, Geosciences and Engineering Sciences, already strongly involved in the problems of risk assessment and management, in particular natural risks. The PhD thesis of Meryem Bousebata is one of the eleven PhDs funded by this project.

- The NeuroCoG project aims at understanding the biological, neurophysiological and functional bases of behavioral and cognitive processes in normal and pathological conditions, from cells to networks and from individual to social cognition. No decisive progress can be achieved in this area without an aspiring interdisciplinary approach. The interdisciplinary ambition of NeuroCoG is particularly strong, bringing together the best scientists, engineers and clinicians at the crossroads of experimental and life sciences, human and social sciences and information and communication sciences, to answer major questions on the workings of the brain and of cognition. One of the work package entitled InnobioPark is dedicated to Parkinson's Disease. The PhD thesis of Veronica Munoz Ramirez is one of the three PhDs in this work package.

- The Grenoble Alpes Data Institute aims at undertaking groundbreaking interdisciplinary research focusing on how data change science and society. It combines three fields of data-related research in a unique way: data science applied to spatial and environmental sciences, biology, and health sciences; data-driven research as a major tool in Social Sciences and Humanities; and studies about data governance, security and the protection of data and privacy. In this context, a 2-year multi-disciplinary projects has been granted in November 2018 to Mistis in collaboration with the Grenoble Institute of Neuroscience. The objective of this project is to develop a statistical learning technique that is able to solve a problem of tracking and analyzing a large population of single molecules. The main difficulties are: 1) the large number of observations to analyse, 2) the noisy nature of the signals, 3) the definition of a quality index to allow the elimination of poor-quality data and false positive signals. We also aim at providing a powerful, well-documented and open-source software, that will be user-friendly for non-specialists.

Also in the context of the Idex associated with the Université Grenoble Alpes, Alexandre Constantin was awarded half a PhD funding from IRS (Initiatives de Recherche Stratégique), 50 keuros.

9.1.2. Competitivity Clusters

The MINALOGIC VISION 4.0 project: MISTIS is involved in a three-year (2016-19) project. The project is led by VI-Technology, a world leader in Automated Optical Inspection (AOI) of a broad range of electronic components. The other partners are the G-Scop Lab in Grenoble and ACTIA company based in Toulouse. Vision 4.0 (in short Vi4.2) is one of the 8 projects labeled by Minalogic, the digital technology competitiveness cluster in Auvergne-Rhône-Alpes, that has been selected for the Industry 4.0 topic in 2016, as part of the 22nd call for projects of the FUI-Régions, for a total budget of the project of 3,4 Meuros.

Today, in the printed circuits boards (PCB) assembly industry, the assembly of electronic cards is a succession of ultra automated steps. Manufacturers, in constant quest for productivity, face sensitive and complex adjustments to reach ever higher levels of quality. Project VI4.2 proposes to build an innovative software solution to facilitate these adjustments, from images and measures obtained in automatic optical inspection (AOI). The idea is - from a centralized station for all the assembly line devices - to analyze and model the defects finely, to adjust each automatic machine, and to configure the interconnection logic between them to improve the quality. Transmitted information is essentially of statistical nature and the role of sc mistis is to identify which statistical methods might be useful to exploit at best the large amount of data registered by AOI machines. Preliminary experiments and results on the Solder Paste Inspection (SPI) step, at the beginning of the assembly line, helped determining candidate variables and measurements to identify future defects and to discriminate between them. More generally, the idea is to analyze two databases at both ends (SPI and Component Inspection) of the assembly process so as to improve our understanding of interactions in

the assembly process, find out correlations between defects and physical measures and generate accordingly proactive alarms so as to detect as early as possible departures from normality.

9.1.3. Networks

MSTGA and AIGM INRA (French National Institute for Agricultural Research) networks: F. Forbes and J.B Durand are members of the INRA network called AIGM (ex MSTGA) network since 2006, http://carlit.toulouse.inra.fr/AIGM, on Algorithmic issues for Inference in Graphical Models. It is funded by INRA MIA and RNSC/ISC Paris. This network gathers researchers from different disciplines. MISTIS co-organized and hosted 2 of the network meetings in 2008 and 2015 in Grenoble. Thibaud Rahier and Brice Olivier participated as speakers to meetings of the network in 2018.

9.2. International Initiatives

9.2.1. Inria International Labs

International Laboratory for Research in Computer Science and Applied Mathematics
Associate Team involved in the International Lab:

9.2.1.1. SIMERG2E

Title: Statistical Inference for the Management of Extreme Risks, Genetics and Global Epidemiology

International Partner:

UGB (Senegal) - Abdou Kâ Diongue

Start year: 2018

See also: http://mistis.inrialpes.fr/simerge

SIMERG2E is built on the same two research themes as SIMERGE, with some adaptations to new applications: 1) Spatial extremes, application to management of extreme risks. We address the definition of new risk measures, the study of their properties in case of extreme events and their estimation from data and covariate information. Our goal is to obtain estimators accounting for possible variability, both in terms of space and time, which is of prime importance in many hydrological, agricultural and energy contexts. 2) Classification, application to genetics and global epidemiology. We address the challenge to build statistical models in order to test association between diseases and human host genetics in a context of genome-wide screening. Adequate models should allow to handle complexity in genomic data (correlation between genetic markers, high dimensionality) and additional statistical issues present in data collected from a family-based longitudinal survey (non-independence between individuals due to familial relationship and non-independence within individuals due to repeated measurements on a same person over time).

9.2.2. Inria International Partners

9.2.2.1. Informal International Partners

The context of our research is also the collaboration between MISTIS and a number of international partners such as the statistics department of University of Michigan, in Ann Arbor, USA, the statistics department of McGill University in Montreal, Canada, Université Gaston Berger in Senegal and Universities of Melbourne and Brisbane in Australia.

The main active international collaborations in 2018 are with:
- G. Stupfler, Nottingham University, UK.
- K. Qin, H. Nguyen and Kerrie Mengersen, D. Wraith resp. from Swinburne University and La Trobe university in Melbourne, Australia and Queensland University of Technology in Brisbane, Australia.
- E. Deme and S. Sylla from Gaston Berger university and IRD in Senegal.
- M. Houle from National Institute of Informatics, Tokyo, Japan.
- N. Wang and C-C. Tu from University of Michigan, Ann Arbor, USA.
- R. Steele, from McGill university, Montreal, Canada.
- Guillaume Kon Kam King, Stefano Favaro, Pierpaolo De Blasi, Collegio Carlo Alberto, Turin, Italy.
- Igor Prünster, Antonio Lijoi, and Riccardo Corradin Bocconi University, Milan, Italy.
- Bernardo Nipoti, Trinity College Dublin, Ireland.
- Yeh Whye Teh, Oxford University and DeepMind, UK.
- Stephen Walker, University of Texas at Austin, USA.

9.3. International Research Visitors

9.3.1. Visits of International Scientists
- Hien Nguyen, researcher at La Trobe University in Melbourne visited for a month in October 2018.
- Eric Marchand Professor at University of Sherbrook Canada, visited from March to June 2018.
- Riccardo Corradin, PhD student at Bocconi University, Milan, Italy visited for a month in March 2018.
- Aboubacrène Ag Ahmad, PhD student at Univ. Gaston Berger, Senegal visited from September 2018 until November 2018.

9.3.1.1. Internships

Caroline Lawless from University College Dublin visited for 2 months as part of her internship.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair
- Florence Forbes, Stéphane Girard and Julyan Arbel organized the two-day workshop Bayesian learning theory for complex data modelling, on September 6-7 2018.

10.1.1.2. Member of the Organizing Committees
- Florence Forbes was a member of the scientific committee of the 50th journées de statistique of Société Francaise de Statistique (JDS 2018) organized in Saclay.
- Julyan Arbel co-organized the two-day workshop entitled Workshop sur la dynamique des communautés sur Twitter en période électorale : analyse par graphes aléatoires workshop on random graphs in Grenoble on April 26-27 2018.
- Julyan Arbel co-organized with Richard Nickl, Cambridge University, a session entitled Bayesian nonparametrics for stochastic processes at International Society for Bayesian Analysis (ISBA) World Meeting 2018 in Edinburgh.
- Jean-Baptiste Durand co-organized a three-day workshop on Models and Analysis of Eye Movements in Grenoble on June 6-8 2018 (https://eyemovements.sciencesconf.org/).

Seminars organization
- MISTIS participates in the weekly statistical seminar of Grenoble. Several lecturers have been invited in this context.
- Florence Forbes, Julyan Arbel and Marta Crispino are co-organizing a monthly reading group on Bayesian statistics.

10.1.2. Scientific Events Selection

10.1.2.1. Reviewer

In 2018, Florence Forbes, Stéphane Girard and Julyan Arbel have been a reviewer for *Journées de la Statistique* (JDS 2018). Additionally,

- In 2018, Julyan Arbel has been a reviewer for
    - Statistics Conferences: *Bayesian Young Statisticians Meeting proceedings* (BAYSM),
    - Machine Learning Conferences: *Conference on Neural Information Processing Systems* (NIPS), *International Conference on Learning Representations* (ICLR), *Symposium on Advances in Approximate Bayesian Inference* (AABI).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Stéphane Girard is Associate Editor of the *Statistics and Computing* journal since 2012 and Associate Editor of the *Journal of Multivariate Analysis* since 2016. He is also member of the Advisory Board of the *Dependence Modelling* journal since December 2014.
- Florence Forbes is Associate Editor of the journal Frontiers in ICT: Computer Image Analysis since its creation in Sept. 2014. She is also Associate Editor of the *Computational Statistics and Data Analysis* journal since May 2018.
- Julyan Arbel is Associate Editor of the *Bayesian Analysis* (BA) journal.

10.1.3.2. Reviewer - Reviewing Activities

- In 2018, Florence Forbes has been a reviewer for *Ecological Modelling* journal.
- In 2018, Stéphane Girard has been a reviewer for *Annals of the Institute of Statistical Mathematics, Statistics & Risk Modeling, Communications in Statistics - Theory and Methods, Extremes.*
- In 2018, Jean-Baptiste Durand has been a reviewer for *Behavior Research Methods* (BRM) and a guest editor for *PLOS Computational Biology* (PLOS Comput. Biol.).
- In 2018, Julyan Arbel has been a reviewer for: *Annals of Statistics* (AoS), *Bayesian Analysis* (BA), *Brazilian Journal of Probability and Statistics* (BJPS), *Computational Statistics & Data Analysis* (CSDA), *Electronic Journal of Statistics* (EJS), *Journal of Nonparametric Statistics* (JNS), *Scandinavian Journal of Statistics* (SJS), *Statistics and Probability Letters* (SPL).

10.1.4. Invited Talks

Florence Forbes has been invited to give talks at the following seminars and conferences:

- Data Science Seminar Series, December 2018 (Link) .
- 11th International Conference of Computational and Methodological Statistics (CMStat), University of Pisa, Italy, December 14-16, [60].
- NeuroCog Seminar Series, October 2018.
- La Trobe-Kyushu Joint Seminar on Mathematics for Industry, Melbourne, October 2018 (Link).
- Joint Statistical Meeting of the American Statistical Association, Vancouver Canada, July [29]
- Workshop on Bayesian nonparametrics, Bordeaux, France, July 2-4, [30]

Julyan Arbel has been invited to give talks at the following seminars and conferences:

- 11th International Conference of Computational and Methodological Statistics (CMStat), University of Pisa, Italy, December 14-16. Invited talk: Some distributional properties of Bayesian neural networks.
- Workshop on Bayesian nonparametrics, Bordeaux, France, July 2-4. Invited talk. Some distributional properties of Bayesian neural networks.
- Olympiades Académiques de Mathématiques, Grenoble. Talk: The mathematics of artificial intelligence.
- Trinity College Statistics Seminar, Dublin, Ireland, May 9. Invited talk: Bayesian graphs and neural networks.
- Journées statistiques de Rochebrune, Megève, France (26-30 March). Invited course: An introduction to Bayesian nonparametric statistics.
- R User group in Grenoble, France, February 8, 2018. Talk (with Alexis Arnaud): Good coding practice, coding style and R packages.
- Workshop on Statistical Methods for Post Genomic Data (SMPGD), Université de Montpellier, France, 11-12 January 2018. Invited talk: A Bayesian Nonparametric Approach to Ecological Risk Assessment.

Among the conferences listed in Section 10, [42], [43], [31], [33] were invited talks.

10.1.5. Scientific Expertise

Florence Forbes is Scientific Advisor since March 2015 for the Pixyl company.

10.1.6. Research Administration

- S. Girard is a member of the "Comité des Emplois Scientifiques" at Inria Grenoble Rhône-Alpes since 2015.
- Since 2015, S. Girard is a member of the INRA committee (CSS MBIA) in charge of evaluating INRA researchers once a year in the MBIA dept of INRA.
- S. Girard has been a reviewer of research projects for the Research Foundation Flanders (FWO), Belgium.
- Florence Forbes is a member of the "Comité Développement Technologique" for software development projects at Inria Grenoble Rhône-Alpes since 2015.
- Florence Forbes is a member of the "Comite d'organisation stratégique" of Inria Grenoble Rhône-Alpes since 2017.
- Florence Forbes is a member of the Executive Committee of the Grenoble data institute.
- Florence Forbes has been a member of the Selection committee for assistant professors at ENS Paris and a member of the Inria admission committee of junior researchers (CRCN) in June 2018.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master : Stéphane Girard, *Statistique Inférentielle Avancée*, 18 ETD, M1 level, Ensimag. Grenoble-INP, France.
- Master : Stéphane Girard, *Data analysis, linear models and ANOVA*, 18 ETD, M1 level, MSIAM. UGA, France.
- Master and PhD course: Julyan Arbel, Bayesian statistics, Ensimag, Université Grenoble Alpes (UGA), 25 ETD.
- Master and PhD course: Julyan Arbel, Bayesian nonparametric statistics, Master Mathématiques Apprentissage et Sciences Humaines (M*A*S*H), Université Paris-Dauphine, 25 ETD.

- Master: Jean-Baptiste Durand, *Statistics and probability*, 192H, M1 and M2 levels, Ensimag Grenoble INP, France. Head of the MSIAM M2 program, in charge of the data science track.
- Jean-Baptiste Durand is a faculty member at Ensimag, Grenoble INP.

10.2.2. Supervision

- PhD defended: Clément Albert "Estimation des limites d'extrapolation par les lois de valeurs extrêmes. Application à des données environnementales", December 2018, Stéphane Girard, Université Grenoble Alpes.
- PhD defended: Thibaud Rahier "Réseaux Bayesiens pour la fusion de données statiques et temporelles", December 2018, Florence Forbes and Stéphane Girard, Université Grenoble Alpes.
- PhD defended: Pierre-Antoine Rodesch "Méthodes statistiques de reconstruction tomographique spectrale pour des systèmes à détection spectrométrique de rayons X", October 9, 2018, Florence Forbes, Université Grenoble Alpes.
- PhD defended: Alexis Arnaud "Analyse statistique d'IRM quantitatives par modèles de mélange : Application à la localisation et la caractérisation de tumeurs cérébrales", October 24, 2018, Florence Forbes and E. Barbier, Université Grenoble Alpes.
- PhD in progress: Karina Ashurbekova, "Robust Graphical Models" Florence Forbes and Sophie Achard, Université Grenoble Alpes, started on October 2016.
- PhD in progress: Veronica Munoz,"Extraction de signatures dans les données IRM de patients parkinsoniens de novo", Florence Forbes and Michel Dojat, Université Grenoble Alpes, started on October 2017.
- PhD in progress: Fabien Boux,"Développement de méthodes statistiques pour l'imagerie IRM fingerprinting", Florence Forbes and Emmanuel Barbier, Université Grenoble Alpes, started on October 2017.
- PhD in progress: Benoit Kugler, "Massive hyperspectral images analysis by inverse regression of physical models", Florence Forbes and Sylvain Douté, Université Grenoble Alpes, started on October 2018.
- PhD in progress: Chun-Chen Tu,"Gaussian mixture sub-clustering/reduction refinement of Non-linear high-to-low dimensional mapping", Florence Forbes and Naisyin Wang, University of Michigan, Ann Arbor.
- PhD in progress: Mariia Vladimirova, "Prior specification for Bayesian deep learning models and regularization implications", started on October 2018, Julyan Arbel and Jakob Verbeek.
- PhD in progress: Brice Olivier, "Joint analysis of eye-movements and EEGs using coupled hidden Markov and topic models", started on October 2015, Jean-Baptiste Durand and Anne Guérin-Dugué (Université Grenoble Alpes).
- PhD in progress: Aboubacrène Ag Ahmad "*A new location-scale model for heavy-tailed distributions*", started on September 2016, Séphane Girard and Alio Diop (Université Gaston Berger, Sénégal).
- PhD in progress: Meryem Bousebata "*Bayesian estimation of extreme risk measures: Implication for the insurance of natural disasters*", started on October 2018, Séphane Girard and Geffroy Enjolras (Université Grenoble Alpes).
- PhD in progress: Alexandre Constantin "*Analyse de séries temporelles massives d'images satellitaires: Applications à la cartographie des écosystèmes*", started on November 2018, Séphane Girard and Mathieu Fauvel (Université Grenoble Alpes).

10.2.3. Juries

- Julyan Arbel has been reviewer for the PhD thesis of Ilaria Bianchini "Modeling and computational aspects of dependent completely random measures in Bayesian nonparametric statistics", Politecnico di Milano, Italy.

- Stéphane Girard has been reviewer for the PhD thesis of Mor Absa Loum, "*Modèle de mélange et modèles linéaires généralisés, application aux données de co-infection*", Univ. Paris-Saclay, France, et Gaston Berger, Sénégal.
- Stéphane Girard has been a member of the PhD committee of Antoine Usseglio Carleve, "*Estimation de mesures de risque pour les distributions elliptiques conditionnées*", Univ. Lyon, France.
- Florence Forbes has been reviewer for the PhD thesis of Amy Chan, University of Queensland, Brisbane and for the HDR thesis of Emilie Lebarbier, agroParisTech.
- Florence Forbes has been a member of the PhD committee of Israel Gebru Inria Grenoble, Marine Roux Gipsa-Lab, Grenoble and Jessica Sodjo, Bordeaux University.

10.3. Popularization

Florence Forbes was a speaker at the Paris Biotech Sante Forum on AI in life sciences, in November 2018, (Program).

10.3.1. Articles and contents

S. Girard and C. Albert have given an interview "When statistics help to predict disasters" for Citizen press: https://www.inria.fr/en/centre/grenoble/news/statistics-and-disasters

11. Bibliography

Major publications by the team in recent years

[1] C. AMBLARD, S. GIRARD. *Estimation procedures for a semiparametric family of bivariate copulas*, in "Journal of Computational and Graphical Statistics", 2005, vol. 14, n° 2, pp. 1–15

[2] J. BLANCHET, F. FORBES. *Triplet Markov fields for the supervised classification of complex structure data*, in "IEEE trans. on Pattern Analyis and Machine Intelligence", 2008, vol. 30(6), pp. 1055–1067

[3] C. BOUVEYRON, S. GIRARD, C. SCHMID. *High dimensional data clustering*, in "Computational Statistics and Data Analysis", 2007, vol. 52, pp. 502–519

[4] C. BOUVEYRON, S. GIRARD, C. SCHMID. *High dimensional discriminant analysis*, in "Communication in Statistics - Theory and Methods", 2007, vol. 36, n° 14

[5] L. CHAARI, T. VINCENT, F. FORBES, M. DOJAT, P. CIUCIU. *Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach*, in "IEEE Transactions on Medical Imaging", May 2013, vol. 32, n° 5, pp. 821-837 [*DOI :* 10.1109/TMI.2012.2225636], http://hal.inria.fr/inserm-00753873

[6] A. DELEFORGE, F. FORBES, R. HORAUD. *High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables*, in "Statistics and Computing", February 2014 [*DOI :* 10.1007/s11222-014-9461-5], https://hal.inria.fr/hal-00863468

[7] F. FORBES, G. FORT. *Combining Monte Carlo and Mean field like methods for inference in hidden Markov Random Fields*, in "IEEE trans. Image Processing", 2007, vol. 16, n° 3, pp. 824-837

[8] F. FORBES, D. WRAITH. *A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering*, in "Statistics and Computing", November 2014, vol. 24, n° 6, pp. 971-984 [*DOI :* 10.1007/s11222-013-9414-4], https://hal.inria.fr/hal-00823451

[9] S. GIRARD. *A Hill type estimate of the Weibull tail-coefficient*, in "Communication in Statistics - Theory and Methods", 2004, vol. 33, n° 2, pp. 205–234

[10] S. GIRARD, P. JACOB. *Extreme values and Haar series estimates of point process boundaries*, in "Scandinavian Journal of Statistics", 2003, vol. 30, n° 2, pp. 369–384

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] C. ALBERT. *Estimation of extrapolation limits based on extreme-value distributions. Application to environmental data*, Communauté Universite Grenoble Alpes, December 2018, https://hal.archives-ouvertes.fr/tel-01971408

[12] A. ARNAUD. *Statistical analysis of quantative MRI based on mixture models : Application to the localization and characterization of brain tumors*, Communauté Université Grenoble-Alpes, October 2018, https://hal.archives-ouvertes.fr/tel-01971217

[13] T. RAHIER. *Bayesian networks for static and temporal data fusion*, Communauté Université Grenoble-Alpes, December 2018, https://hal.archives-ouvertes.fr/tel-01971371

Articles in International Peer-Reviewed Journals

[14] L. AMSALEG, O. CHELLY, T. FURON, S. GIRARD, M. E. HOULE, K.-I. KAWARABAYASHI, M. NETT. *Extreme-value-theoretic estimation of local intrinsic dimensionality*, in "Data Mining and Knowledge Discovery", November 2018, vol. 32, n° 6, pp. 1768–1805 [*DOI :* 10.1007/S10618-018-0578-6], https://hal.archives-ouvertes.fr/hal-01864580

[15] J. ARBEL, P. DE BLASI, I. PRÜNSTER. *Stochastic approximations to the Pitman-Yor process*, in "Bayesian Analysis", 2018, pp. 1-19 [*DOI :* 10.1214/18-BA1127], https://hal.archives-ouvertes.fr/hal-01950654

[16] A. ARNAUD, F. FORBES, N. COQUERY, N. COLLOMB, B. L. LEMASSON, E. L. BARBIER. *Fully Automatic Lesion Localization and Characterization: Application to Brain Tumors Using Multiparametric Quantitative MRI Data*, in "IEEE Transactions on Medical Imaging", July 2018, vol. 37, n° 7, pp. 1678-1689 [*DOI :* 10.1109/TMI.2018.2794918], https://hal.archives-ouvertes.fr/hal-01545548

[17] O. COMMOWICK, A. ISTACE, M. KAIN, B. LAURENT, F. LERAY, M. SIMON, S. CAMARASU-POP, P. GIRARD, R. AMELI, J.-C. FERRÉ, A. KERBRAT, T. TOURDIAS, F. CERVENANSKY, T. GLATARD, J. BEAUMONT, S. DOYLE, F. FORBES, J. KNIGHT, A. KHADEMI, A. MAHBOD, C. WANG, R. MCKINLEY, F. WAGNER, J. MUSCHELLI, E. SWEENEY, E. ROURA, X. LLADÓ, M. M. SANTOS, W. P. SANTOS, A. G. SILVA-FILHO, X. TOMAS-FERNANDEZ, H. URIEN, I. BLOCH, S. VALVERDE, M. CABEZAS, F. J. VERA-OLMOS, N. MALPICA, C. R. G. GUTTMANN, S. VUKUSIC, G. EDAN, M. DOJAT, M. STYNER, S. K. WARFIELD, F. COTTON, C. BARILLOT. *Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure*, in "Scientific Reports", September 2018, vol. 8, 13650 p. [*DOI :* 10.1038/S41598-018-31911-7], https://www.hal.inserm.fr/inserm-01847873

[18] A. DAOUIA, S. GIRARD, G. STUPFLER. *Estimation of Tail Risk based on Extreme Expectiles*, in "Journal of the Royal Statistical Society: Series B", March 2018, vol. 80, n° 2, pp. 263–292 [*DOI :* 10.1111/RSSB.12254], https://hal.archives-ouvertes.fr/hal-01142130

[19] A. DAOUIA, S. GIRARD, G. STUPFLER. *Extreme M-quantiles as risk measures: From L1 to Lp optimization*, in "Bernoulli", 2018, forthcoming, https://hal.inria.fr/hal-01585215

[20] J. EL METHNI, L. GARDES, S. GIRARD. *Kernel estimation of extreme regression risk measures*, in "Electronic journal of statistics ", 2018, vol. 12, n° 1, pp. 359–398, https://hal.inria.fr/hal-01393519

[21] M. GARBEZ, R. SYMONEAUX, É. BELIN, Y. CARAGLIO, Y. CHÉNÉ, N. DONES, J.-B. DURAND, G. HUNAULT, D. RELION, M. SIGOGNE, D. ROUSSEAU, G. GALOPIN. *Ornamental plants architectural characteristics in relation to visual sensory attributes: a new approach on the rose bush for objective evaluation of the visual quality*, in "European Journal of Horticultural Science", 2018, vol. 83, n° 3, pp. 187-201 [*DOI :* 10.17660/EJHS.2018/83.3.8], https://hal.archives-ouvertes.fr/hal-01831318

[22] S. GIRARD. *Transformation of a copula using the associated co-copula*, in "Dependence Modeling", 2018, vol. 6, pp. 298-308, https://hal.inria.fr/hal-01881969

[23] E. PERTHAME, F. FORBES, A. DELEFORGE. *Inverse regression approach to robust nonlinear high-to-low dimensional mapping*, in "Journal of Multivariate Analysis", January 2018, vol. 163, pp. 1 - 14 [*DOI :* 10.1016/J.JMVA.2017.09.009], https://hal.inria.fr/hal-01347455

Invited Conferences

[24] J. ARBEL. *A Bayesian Nonparametric Approach to Ecological Risk Assessment*, in "SMPGD 2018 - Workshop on Statistical Methods for Post Genomic Data", Montpellier, France, January 2018, https://hal.archives-ouvertes.fr/hal-01950669

[25] J. ARBEL. *An introduction to Bayesian nonparametric statistics*, in "2018 - Journées statistiques de Roche-brune", Megève, France, March 2018, https://hal.archives-ouvertes.fr/hal-01950668

[26] J. ARBEL. *Bayesian neural network priors at the level of units*, in "CMStat 2018 - 11th International Conference of Computational and Methodological Statistics", Pisa, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01950661

[27] J. ARBEL. *Some distributional properties of Bayesian neural networks*, in "Workshop on Bayesian nonparametrics", Bordeaux, France, July 2018, https://hal.archives-ouvertes.fr/hal-01950667

[28] F. FORBES, A. ARNAUD, R. STEELE, B. LEMASSON, E. L. BARBIER. *Bayesian mixtures of multiple scale distributions*, in "CMStatistics 2018 - 11th International Conference of the ERCIM WG on Computational and Methodological Statistics", Pisa, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01941682

[29] F. FORBES, H. LU, J. ARBEL. *Non parametric Bayesian priors for hidden Markov random fields*, in "JSM 2018 - Joint Statistical Meeting", Vancouver, Canada, JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association, July 2018, pp. 1-38, https://hal.archives-ouvertes.fr/hal-01941679

[30] F. FORBES, H. LU, J. ARBEL. *Non parametric Bayesian priors for hidden Markov random fields: application to image segmentation*, in "BNPSI 2018 : Workshop on Bayesian non parametrics for signal and image processing", Bordeaux, France, July 2018, https://hal.archives-ouvertes.fr/hal-01941687

[31] S. GIRARD. *Estimation of extreme regression risk measures*, in "2018 - Workshop Rare Events, Extremes and Machine Learning", Paris, France, May 2018, https://hal.inria.fr/hal-01800772

[32] T. RAHIER, S. MARIÉ, S. GIRARD, F. FORBES. *Screening strong pairwise relationships for fast Bayesian network structure learning*, in "2nd Italian-French Statistics Seminar - IFSS", Grenoble, France, September 2018, https://hal.archives-ouvertes.fr/hal-01941685

[33] G. STUPFLER, S. GIRARD. *Estimation of high-dimensional extreme conditional expectiles*, in "CMStatistics 2018 - 11th International Conference of the ERCIM WG on Computing and Statistics", Pisa, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01942210

International Conferences with Proceedings

[34] R. AZAÏS, J.-B. DURAND, C. GODIN. *Approximation of trees by self-nested trees*, in "ALENEX 2019 - Algorithm Engineering and Experiments", San Diego, United States, January 2019, pp. 1-24, https://arxiv.org/abs/1810.10860 [*DOI :* 10.10860], https://hal.archives-ouvertes.fr/hal-01294013

[35] F. BOUX, F. FORBES, J. ARBEL, E. L. BARBIER. *Dictionary-Free MR Fingerprinting Parameter Estimation Via Inverse Regression*, in "Joint Annual Meeting ISMRM-ESMRMB 2018", Paris, France, Proceedings of Joint Annual Meeting ISMRM-ESMRMB 2018, June 2018, pp. 1-2, https://hal.archives-ouvertes.fr/hal-01941630

[36] V. MUÑOZ RAMÍREZ, F. FORBES, J. ARBEL, A. ARNAUD, M. DOJAT. *Quantitative MRI characterization of brain abnormalities in de novo Parkinsonian patients*, in "IEEE International Symposium on Biomedical Imaging", Venice, Italy, Proceedings of IEEE International Symposium on Biomedical Imaging, April 2019, https://hal.archives-ouvertes.fr/hal-01970682

[37] P.-A. RODESCH, V. REBUFFEL, C. FOURNIER, F. FORBES, L. VERGER. *Spectral CT reconstruction with an explicit photon-counting detector model: a " one-step " approach*, in "SPIE Medical Imaging", Houston, United States, February 2018, vol. 10573, pp. 1-4 [*DOI :* 10.1117/12.2285792], https://hal.inria.fr/hal-01652017

National Conferences with Proceedings

[38] C. ALBERT, A. DUTFOY, L. GARDES, S. GIRARD. *Un nouvel estimateur des quantiles extrêmes basé sur le modèle "log Weibull-tail" généralisé*, in "SFDS 2018 - 50èmes Journées de Statistique", Saclay, France, May 2018, pp. 1-6, https://hal.inria.fr/hal-01807672

[39] K. ASHURBEKOVA, S. ACHARD, F. FORBES. *Robust structure learning using multivariate T-distributions*, in "50e Journées de la Statistique de la SFdS", Saclay, France, May 2018, pp. 1-6, https://hal.archives-ouvertes.fr/hal-01941643

[40] H. LU, J. ARBEL, F. FORBES. *Bayesian Nonparametric Priors for Hidden Markov Random Fields*, in "50e Journées de la Statistique de la SFdS", Saclay, France, Actes 50e Journées de la Statistique de la SFdS, May 2018, pp. 1-5, https://hal.archives-ouvertes.fr/hal-01941638

[41] T. RAHIER, S. MARIÉ, S. GIRARD, F. FORBES. *Fast Bayesian Network Structure Learning using Quasi-Determinism Screening*, in "9èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes", Toulouse, France, Actes des 9èmes Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes, June 2018, https://hal.archives-ouvertes.fr/hal-01941683

Conferences without Proceedings

[42] C. ALBERT, A. DUTFOY, S. GIRARD. *Extrapolation limits of extreme-value methods for return-levels estimation*, in "EGU General Assembly", Vienna, Austria, April 2018, https://hal.inria.fr/hal-01773660

[43] J. EL METHNI, L. GARDES, S. GIRARD. *Kernel estimation of extreme regression risk measures*, in "ICOR 2018 - 13th International Conference on Operations Research", La Havane, Cuba, March 2018, 1 p. , https://hal.archives-ouvertes.fr/hal-01745322

[44] M. VLADIMIROVA, J. ARBEL, P. MESEJO. *Bayesian neural network priors at the level of units*, in "AABI 2018 - 1st Symposium on Advances in Approximate Bayesian Inference", Montréal, Canada, December 2018, pp. 1-6, https://hal.archives-ouvertes.fr/hal-01950659

[45] M. VLADIMIROVA, J. ARBEL, P. MESEJO. *Bayesian neural networks become heavier-tailed with depth*, in "NeurIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, pp. 1-7, https://hal.archives-ouvertes.fr/hal-01950658

[46] F. ZHENG, M. JALBERT, F. FORBES, S. BONNET, A. WOJTUSCISZYN, S. LABLANCHE, P.-Y. BENHAMOU. *Caractérisation de la variabilité glycémique journalière chez le patient avec diabète de type 1*, in "SFD 2018 - Congrès annuel de la Société Francophone du Diabète", Marseille, France, March 2019, https://hal.archives-ouvertes.fr/hal-01971621

Scientific Books (or Scientific Book chapters)

[47] J. ARBEL. *Clustering Milky Way's Globulars: a Bayesian Nonparametric Approach*, in "Statistics for Astrophysics: Bayesian Methodology", 2018, pp. 113-137, https://hal.archives-ouvertes.fr/hal-01950656

[48] M. CLAUSEL, J.-B. DURAND. *Modèles génératifs*, in "Data Science. Cours et exercices", Eyrolles, August 2018, https://hal.inria.fr/hal-01961722

[49] F. FORBES. *Mixture Models for Image Analysis*, in "Handbook of Mixture Analysis", CRC press, December 2018, https://hal.archives-ouvertes.fr/hal-01970681

[50] C. MAGGIA, T. MISTRAL, S. DOYLE, F. FORBES, A. KRAINIK, D. GALANAUD, E. SCHMITT, S. KREMER, I. TROPRÈS, E. L. BARBIER, J.-F. PAYEN, M. DOJAT. *Traumatic Brain Lesion Quantification based on Mean Diffusivity Changes*, in "Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, BrainLes (MICCAI), ed Crimi Aea (Springer International Publishing AG)", 2018, https://hal.archives-ouvertes.fr/hal-01704679

[51] K. K. MENGERSEN, E. DUNCAN, J. ARBEL, C. ALSTON-KNOX, N. WHITE. *Applications in Industry*, in "Handbook of mixture analysis", CRC press, 2018, https://hal.archives-ouvertes.fr/hal-01963798

Other Publications

[52] A. AG AHMAD, A. DIOP, E. H. DEME, S. GIRARD. *A new location-scale model for conditional heavy-tailed distributions*, September 2018, IFSS 2018 - 2nd Italian-French Statistics Seminar, Poster, https://hal.archives-ouvertes.fr/hal-01942204

[53] C. ALBERT, A. DUTFOY, L. GARDES, S. GIRARD. *An extreme quantile estimator for the log-generalized Weibull-tail model*, September 2018, working paper or preprint, https://hal.inria.fr/hal-01783929

[54] C. ALBERT, A. DUTFOY, S. GIRARD. *Asymptotic behavior of the extrapolation error associated with the estimation of extreme quantiles*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01692544

[55] J. ARBEL. *Bayesian Nonparametric Mixtures Why and How?*, September 2018, IFSS 2018 - 2nd Italian-French Statistics Seminar, Poster, https://hal.archives-ouvertes.fr/hal-01950664

[56] J. ARBEL, R. CORRADIN, M. LEWANDOWSKI. *Comment on Article by Wade and Ghahramani*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01950655

[57] J. ARBEL, R. CORRADIN, B. NIPOTI. *Dirichlet process mixtures under affine transformations of the data*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01950652

[58] J. ARBEL, M. CRISPINO, S. GIRARD. *Dependence properties and Bayesian inference for asymmetric multivariate copulas*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01963975

[59] J. ARBEL, S. FAVARO. *Approximating predictive probabilities of Gibbs-type priors*, January 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01693333

[60] A. ARNAUD, F. FORBES, R. STEELE, B. LEMASSON, E. L. BARBIER. *Bayesian mixtures of multiple scale distributions*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01953393

[61] P. BRUEL, S. QUINITO MASNADA, B. VIDEAU, A. LEGRAND, J.-M. VINCENT, A. GOLDMAN. *Autotuning under Tight Budget Constraints: A Transparent Design of Experiments Approach*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01953287

[62] A. DAOUIA, S. GIRARD, G. STUPFLER. *ExpectHill estimation, extreme risk and heavy tails*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01856212

[63] A. DAOUIA, S. GIRARD, G. STUPFLER. *Tail expectile process and risk assessment*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01744505

[64] D. FRAIX-BURNET, S. GIRARD, J. ARBEL, J.-B. MARQUETTE. *Statistics for Astrophysics: Bayesian Methodology*, 2018, notice, https://hal.archives-ouvertes.fr/hal-01952759

[65] L. GARDES, S. GIRARD, G. STUPFLER. *Beyond tail median and conditional tail expectation: extreme risk estimation using tail L_p-optimisation*, March 2018, working paper or preprint, https://hal.inria.fr/hal-01726328

[66] C. LAWLESS, J. ARBEL. *A simple proof of Pitman-Yor's Chinese restaurant process from its stick-breaking representation*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01950653

[67] C. LAWLESS, J. ARBEL. *Chinese restaurant process from stick-breaking for Pitman-Yor*, September 2018, 1 p. , Bayesian learning theory for complex data modelling Workshop, Poster, https://hal.archives-ouvertes.fr/hal-01950662

[68] H. Lu, J. Arbel, F. Forbes. *Bayesian Nonparametric Priors for Hidden Markov Random Fields: Application to Image Segmentation*, September 2018, 1 p. , IFSS 2018 - 2nd Italian-French Statistics Seminar, Poster, https://hal.archives-ouvertes.fr/hal-01950666

[69] A. Malkova, J. Arbel, M. L. Delle Monache. *DATASAFE: understanding Data Accidents for TrAffic SAFEty Acknowledgments*, September 2018, 1 p. , Bayesian learning theory for complex data modelling Workshop, Poster, https://hal.archives-ouvertes.fr/hal-01950663

[70] O. Marchal, J. Arbel. *Beta and Dirichlet sub-Gaussianity*, September 2018, Bayesian learning theory for complex data modelling Workshop, Poster, https://hal.archives-ouvertes.fr/hal-01950665

[71] T. Rahier, S. Marié, S. Girard, F. Forbes. *Fast Bayesian Network Structure Learning using Quasi-Determinism Screening*, March 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01691217

[72] C.-C. Tu, F. Forbes, B. Lemasson, N. Wang. *Structured Mixture of Linear Mappings in High Dimension* , February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01700053

[73] M. Vladimirova, J. Arbel, P. Mesejo. *Bayesian neural network priors at the level of units*, November 2018, 1 p. , Bayesian Statistics in the Big Data Era, Poster, https://hal.archives-ouvertes.fr/hal-01950660

[74] M. Vladimirova, J. Arbel, P. Mesejo. *Bayesian neural networks increasingly sparsify their units with depth*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01950657

References in notes

[75] C. Bouveyron. *Modélisation et classification des données de grande dimension. Application à l'analyse d'images*, Université Grenoble 1, septembre 2006, http://tel.archives-ouvertes.fr/tel-00109047

[76] P. Embrechts, C. Klüppelberg, T. Mikosh. *Modelling Extremal Events*, Applications of Mathematics, Springer-Verlag, 1997, vol. 33

[77] F. Ferraty, P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, 2006

[78] S. Girard. *Construction et apprentissage statistique de modèles auto-associatifs non-linéaires. Application à l'identification d'objets déformables en radiographie. Modélisation et classification*, Université de Cery-Pontoise, octobre 1996

[79] K. Li. *Sliced inverse regression for dimension reduction*, in "Journal of the American Statistical Association", 1991, vol. 86, pp. 316–327

[80] B. Olivier, J.-B. Durand, A. Guérin-Dugué, M. Clausel. *Eye-tracking data analysis using hidden semi-Markovian models to identify and characterize reading strategies*, in "19th European Conference on Eye Movements (ECM 2017)", Wuppertal, Germany, August 2017, https://hal.inria.fr/hal-01671224