IN PARTNERSHIP WITH:
Institut polytechnique de Grenoble

Université de Grenoble Alpes

Activity Report 2017

Project-Team MISTIS

Modelling and Inference of Complex and Structured Stochastic Systems

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Optimization, machine learning and statistical methods

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
B9.4.5. - Data science
B9.9.1. - Environmental risks

1. Personnel

Research Scientists

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

Gildas Mazo [Inria, Starting Research Position, until Oct 2017]

Faculty Member

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

Post-Doctoral Fellows

Jean-Michel Bécu [Inria]
Hongliang Lu [Inria, from Oct 2017]
Emeline Perthame [Inria, until Jan 2017]

PhD Students

Clément Albert [Inria]
Alexis Arnaud [Univ Joseph Fourier Grenoble, until Sep 2017; Institut polytechnique de Grenoble, from Oct 2017]
Karina Ashurbekova [Univ Grenoble Alpes]
Fabien Boux [Univ Grenoble Alpes, from Sep 2017]
Aina Frau Pascual [Inria, until Mar 2017]
Veronica Munoz Ramirez [Univ Grenoble Alpes, from Oct 2017]
Brice Olivier [Univ Grenoble Alpes]
Thibaud Rahier [Autre entreprise privée]
Pierre-Antoine Rodesch [CEA]

Technical staff

Jaime Eduardo Arias Almeida [Inria]
Fatima Fofana [Inria, from Oct 2017]
Pascal Rubini [Inria, until Jun 2017]

Interns

Nicolas Allemonière [Inria, from Mar 2017 until Aug 2017]
Yaroslav Averyanov [Inria, from Feb 2017 until Jun 2017]
Fatima Fofana [Inria, from May 2017 until Sep 2017]
Mariem Hbaieb [Inria, from May 2017 until Jul 2017]

Administrative Assistant

Marion Ponsot [Inria]

Visiting Scientist

Aboubacrène Ahmad [Université Gaston Berger, Sénégal, from Sep 2017 until Oct 2017]

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, Aina Frau Pascual, Stéphane Girard, Julyan Arbel, Gildas Mazo, Jean-Michel Bécu, Hongliang Lu, Emeline Perthame, Fabien Boux, Veronica Munoz Ramirez.

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 multidimensional, 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, Pierre-Antoine Rodesch, Julyan Arbel.

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, Emeline Perthame, Jean-Michel Bécu.

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 [67]. 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 [69] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [66], 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).

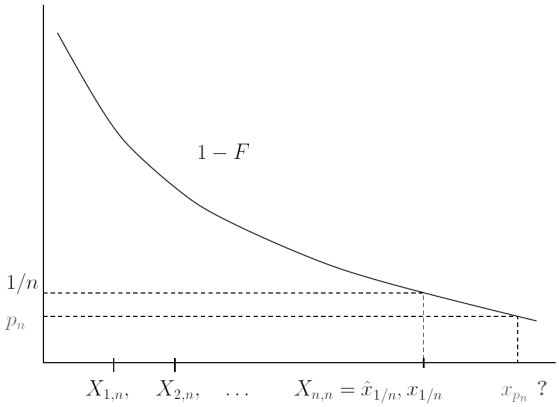

Figure 1. The curve represents the survival function $x \to P(X > x)$. The $1/n$-quantile is estimated by the maximum observation so that $\widehat{x}_{1/n} = X_{n,n}$. As illustrated in the figure, to estimate p_n-quantiles with $p_n < 1/n$, it is necessary to extrapolate beyond the maximum observation.

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.
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:

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

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 distorsions 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 [68]. 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 [65]. 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 [69].

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Aina Frau Pascual, Florence Forbes, Stéphane Girard, Pascal Rubini, Jaime Eduardo Arias Almeida, Pierre-Antoine Rodesch.

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 Laboratoire de Planétologie de Grenoble. We worked on hyperspectral 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: Aina Frau Pascual, Jaime Eduardo Arias Almeida, Alexis Arnaud, Florence Forbes, Stéphane Girard, Emeline Perthame, Jean-Baptiste Durand, Clément Albert, Julyan Arbel, Jean-Michel Bécu, Thibaud Rahier, Brice Olivier, Karina Ashurbekova, Fabien Boux, Veronica Munoz Ramirez.

A third domain of applications concerns biology and medicine. We considered the use of missing data models in epidemiology. We also investigated statistical tools for the analysis of bacterial genomes beyond gene detection. Applications in neurosciences are also considered. In the environmental domain, we considered the modelling of high-impact weather events.

5. Highlights of the Year

5.1. Highlights of the Year

- Veronica Munoz Ramirez supervised by F. Forbes, J. Arbel (MISTIS) and M. Dojat (Grenoble Institute of neuroscience) was granted a PhD grant from the Idex NeuroCoG project. The PhD project is part of a work package, dedicated to Parkinson's Disease (PD), which aims at identifying multidimensional cognitive and neurophysiological biomarkers for early diagnosis, outcome prediction and novel neurorehabilitation methods for de novo PD patients.

- In the context of another Idex project named Grenoble Data Institute, two 2-years multi-disciplinary projects were granted in November 2017 to Mistis in collaboration respectively with Team Necs from Inria and Gipsa-lab (DATASAFE project: understanding Data Accidents for TrAffic SAFEty) and with IPAG and Univ. Paris Sud Orsay (Regression techniques for Massive Mars hyperspectral image analysis from physical model inversion).

6. New Software and Platforms

6.1. BOLD model FIT

KEYWORDS: Functional imaging - FMRI - Health
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

- Authors: Jan M Warnking, Pablo Mesejo Santiago and Florence Forbes
- Contact: Pablo Mesejo Santiago
- URL: https://hal.archives-ouvertes.fr/hal-01221115v2/

6.2. MMST

Mixtures of Multiple Scaled Student T distributions
KEYWORDS: Medical imaging - Brain MRI - Statistics - Health - Robust clustering
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.

- Participants: Alexis Arnaud, Darren Wraith and Florence Forbes
- Contact: Florence Forbes
- URL: http://mistis.inrialpes.fr/realisations.html

6.3. PyHRF

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

- 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
- URL: http://pyhrf.org

6.4. xLLiM

High dimensional locally linear mapping
KEYWORDS: Clustering - Regression

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.

- Participants: Antoine Deleforge, Emeline Perthame and Florence Forbes
- Contact: Florence Forbes
- URL: https://cran.r-project.org/web/packages/xLLiM/index.html

7. New Results

7.1. Mixture models

7.1.1. Robust and collaborative extensions of Sliced Inverse Regression.

Participants: Stéphane Girard, Florence Forbes.

This research theme was supported by a LabEx PERSYVAL-Lab project-team grant.

Joint work with: A. Chiancone and J. Chanussot (Gipsa-lab and Grenoble-INP).

Sliced Inverse Regression (SIR) has been extensively used to reduce the dimension of the predictor space before performing regression. Recently it has been shown that this technique is, not surprisingly, sensitive to noise. Different approaches have thus been proposed to robustify SIR. In [16], we investigate the properties of an inverse problem proposed by R.D. Cook and we show that the framework can be extended to take into account a non-Gaussian noise. Generalized Student distributions are considered and all parameters are estimated via an EM algorithm. The algorithm is outlined and tested comparing the results with different approaches on simulated data. Results on a real dataset show the interest of this technique in presence of outliers.

For further improvement of SIR, in his PhD thesis work, Alessandro Chiancone studied the extension of the SIR method to different sub-populations. The idea is to assume that the dimension reduction subspace is not the same for different clusters of the data [17]. One of the difficulties is that standard Sliced Inverse Regression (SIR) has requirements on the distribution of the predictors that are hard to check since they depend on unobserved variables. It has been shown that, if the distribution of the predictors is elliptical, then these requirements are satisfied. In case of mixture models, the ellipticity is violated and in addition there is no assurance of a single underlying regression model among the different components. Our approach clusters the predictors space to force the condition to hold on each cluster and includes a merging technique to look for different underlying models in the data. A study on simulated data as well as two real applications are provided. It appears that SIR, unsurprisingly, is not able to deal with a mixture of Gaussians involving different underlying models whereas our approach is able to correctly investigate the mixture.

7.1.2. Structured 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. GLLiM is based on a joint modeling of both the responses and covariates, observed or latent. This joint modeling allows for the use of an inverse regression strategy to handle the high dimensionality of the data. Mixtures are used to approximate non-linear associations. GLLiM groups data with similar linear association together. Within the same cluster, the association can be considered as locally linear, which can then be resolved under the classical linear regression setting (see Figure 2(a)). However, when the covariate dimension is much higher than the response dimension, GLLiM may result in erroneous clusters at the low dimension (eg Figure 2 (b)), leading to potentially inaccurate predictions. Specifically, when the clustering is conducted at a high joint dimension, the distance at low dimension between two members of the same cluster (component) could remain large. As a result, a mixture component might contain several sub-clusters and/or outliers, violating the model Gaussian assumption. This results in a model misspecification effect that can seriously impact prediction performance. A natural way to lessen this effect is to increase the number of components in the mixture making each linear mapping even more local. But this also increases the number of parameters to estimate and therefore requires to be done in a parsimonious manner to avoid over-parameterization. In this work, we propose a parsimonious approach which we refer to as Structured Mixture of Gaussian Locally Linear Mapping (SMoGLLiM) to solve the aforementioned problems. It follows a two-layer hierarchical clustering structure where local components are grouped into global components sharing the same high-dimensional noise covariance structure, which effectively reduces the number of parameters of the model. SMoGLLiM also includes a pruning algorithm for eliminating outliers as well as determining an appropriate number of clusters. Moreover, the number of clusters and training outliers determined by SMoGLLiM can be further used by GLLiM for improving prediction performance. As an extension, a subsetting and parallelization techniques are discussed for the efficiency concern. A preliminary version of this work was presented at the American Statistical Association Joint Statistical Meeting in Baltimore USA in July 2017, [35].

7.1.3. 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 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 submitted to ISMRM 2018, International Society for Magnetic Resonance in Medicine.

7.1.4. Semiparametric copula-based clustering

Participants: Florence Forbes, Gildas Mazo, Yaroslav Averyanov.

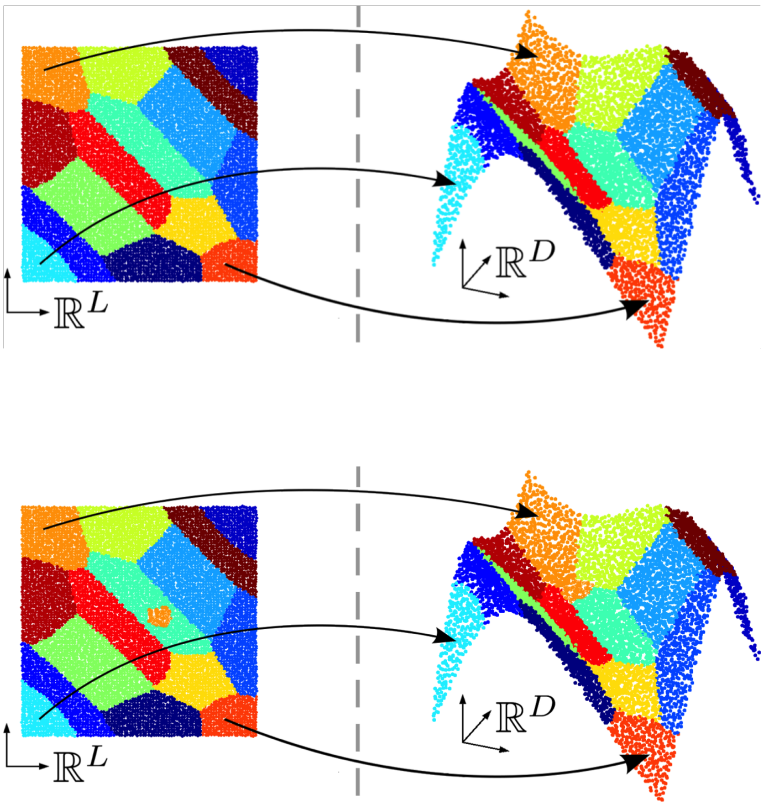

Figure 2. illustration of the GLLiM model: (Top) Non linear relationship approximated as a mixture of locally linear mappings; (Bottom) problematic clustering with a non Gaussian component (orange region) due to unbalanced weights between the high dimensional responses and low dimensional covariates.

Modeling of distributions mixtures has rested on Gaussian distributions and/or a conditional independence hypothesis for a long time. Only recently have researchers begun to construct and study broader generic models without appealing to such hypotheses. Some of these extensions use copulas as a tool to build flexible models, as they permit to model the dependence and the marginal distributions separately. Recently [70], a semiparametric copula-based mixture model has been proposed to cluster continuous data. This semiparametric feature allows for more flexibility and reduces the modelling effort for the practitioner. Nonetheless, these advantages come at the cost of assuming that the clusters do not differ in scale. The aim of the internship of Y. Averyanov was to get rid of this assumption by building a nonparametric estimator which have to satisfy certain moment constraints. The performance of the estimator was tested on simulations and then embedded into an EM-like algorithm framework.

7.1.5. Fully automatic lesion localization and characterization: application to brain tumors using multiparametric quantitative MRI data

Participants: Florence Forbes, Alexis Arnaud.

Joint work with: Emmanuel Barbier, Nora Collomb and Benjamin Lemasson from Grenoble Institute of Neuroscience.

When analyzing brain tumors, two tasks are intrinsically linked, spatial localization and physiological characterization of the lesioned tissues. Automated data-driven solutions exist, based on image segmentation techniques or physiological parameters analysis, but for each task separately, the other being performed manually or with user tuning operations. In this work, the availability of quantitative magnetic resonance (MR) parameters is combined with advanced multivariate statistical tools to design a fully automated method that jointly performs both localization and characterization. Non trivial interactions between relevant physiological parameters are captured thanks to recent generalized Student distributions that provide a larger variety of distributional shapes compared to the more standard Gaussian distributions. Probabilistic mixtures of the former distributions are then considered to account for the different tissue types and potential heterogeneity of lesions. Discriminative multivariate features are extracted from this mixture modelling and turned into individual lesion signatures. The signatures are subsequently pooled together to build a statistical fingerprint model of the different lesion types that captures lesion characteristics while accounting for inter-subject variability. The potential of this generic procedure is demonstrated on a data set of 53 rats, with 36 rats bearing 4 different brain tumors, for which 5 quantitative MR parameters were acquired. This study has been submitted for publication [15].

Analyzing brain tumor tissue composition can then improve the handling of tumor growth and resistance to therapies. We showed on a 6 time point dataset of 8 rats that multiparametric MRI could be exploited via statistical clustering to quantify intra-lesional heterogeneity in space and time. More specifically, MRI can be used to map structural, eg diffusion, as well as functional, eg volume (BVf), vessel size (VSI), oxygen saturation of the tissue (StO2), characteristics. In previous work, these parameters have been analyzed to show the great potential of multiparametric MRI (mpMRI) to monitor combined radio- and chemo-therapies. However, to exploit all the information contained in mpMRI while preserving information about tumor heterogeneity, new methods need to be developed. We demonstrated the ability of clustering analysis applied to longitudinal mpMRI to summarize and quantify intra-lesional heterogeneity during tumor growth. This study showed the interest of a clustering analysis on mpMRI data to monitor the evolution of brain tumor heterogeneity. It highlighted the type of tissue that mostly contributes to tumor development and could be used to refine the evaluation of therapies and to improve tumor prognosis. This work has been presented at ISMRM 2017, International Society for Magnetic Resonance in Medicine [42].

7.1.6. Signature extraction in MR scans for de novo Parkinsonian patients

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

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

This work is part of the cross-disciplinary project NeuroCoG. Parkinson's disease (PD) is characterized by the degeneration of dopaminergic neurons located in the substantia nigra pars compacta (SNc). This leads to well-known motor symptoms associated to Parkinson's disease, rigidity, akinesia and tremor. However, non-motor symptoms also appear. It is of primordial interest to understand these symptoms in order to optimize treatments and diagnose at an early stage the pathology's occurrence. The goal of the PhD work of Veronica Munoz Ramirez is the extraction of specific signatures from MR data of de novo PD patients. We investigated the possibility to use multivariate non-supervised clustering techniques as developed in the PhD thesis of Alexis Arnaud to cluster voxels taking into account interactions between various parameters.

7.1.7. Object-based classification from high resolution satellite image time series with Gaussian mean map kernels

Participant: Stéphane Girard.

Joint work with: C. Bouveyron (Univ. Paris 5), M. Fauvel and M. Lopes (ENSAT Toulouse)

In the PhD work of Charles Bouveyron [65], we proposed new Gaussian models of high dimensional data for classification purposes. We assume that the data live in several groups located in subspaces of lower dimensions. Two different strategies arise:

- the introduction in the model of a dimension reduction constraint for each group
- the use of parsimonious models obtained by imposing to different groups to share the same values of some parameters.

This modelling yielded a supervised classification method called High Dimensional Discriminant Analysis (HDDA)[4]. Some versions of this method have been tested on the supervised classification of objects in images. This approach has been adapted to the unsupervised classification framework, and the related method is named High Dimensional Data Clustering (HDDC)[3]. In the framework of Mailys Lopes PhD, our recent work [22], [23], consists in adapting this work to the classification of grassland management practices using satellite image time series with high spatial resolution. The study area is located in southern France where 52 parcels with three management types were selected. The spectral variability inside the grasslands was taken into account considering that the pixels signal can be modeled by a Gaussian distribution. To measure the similarity between two grasslands, a new kernel is proposed as a second contribution: the a-Gaussian mean kernel. It allows to weight the influence of the covariance matrix when comparing two Gaussian distributions. This kernel is introduced in Support Vector Machine for the supervised classification of grasslands from south-west France. A dense intra-annual multispectral time series of Formosat-2 satellite is used for the classification of grasslands management practices, while an inter-annual NDVI time series of Formosat-2 is used for permanent and temporary grasslands discrimination. Results are compared to other existing pixel- and object-based approaches in terms of classification accuracy and processing time. The proposed method shows to be a good compromise between processing speed and classification accuracy. It can adapt to the classification constraints and it encompasses several similarity measures known in the literature. It is appropriate for the classification of small and heterogeneous objects such as grasslands.

7.2. Semi and non-parametric methods

7.2.1. Robust estimation for extremes

Participants: Clément Albert, Stéphane Girard.

Joint work with: M. Stehlik (Johannes Kepler Universitat Linz, Austria and Universidad de Valparaiso, Chile) and A. Dutfoy (EDF R&D).

In the PhD thesis of Clément Albert (funded by EDF), we study the sensitivity of extreme-value methods to small changes in the data and we investigate their extrapolation ability [36], [37]. To reduce this sensitivity, robust methods are needed and we proposed a novel method of heavy tails estimation based on a transformed score (the t-score). Based on a new score moment method, we derive the t-Hill estimator, which estimates the extreme value index of a distribution function with regularly varying tail. t-Hill estimator is distribution sensitive, thus it differs in e.g. Pareto and log-gamma case. Here, we study both forms of the estimator, i.e. t-Hill and t-lgHill. For both estimators we prove weak consistency in moving average settings as well as the asymptotic normality of t-lgHill estimator in the i.i.d. setting. In cases of contamination with heavier tails than the tail of original sample, t-Hill outperforms several robust tail estimators, especially in small sample situations. A simulation study emphasizes the fact that the level of contamination is playing a crucial role. We illustrate the developed methodology on a small sample data set of stake measurements from Guanaco glacier in Chile. This methodology is adapted to bounded distribution tails in [26] with an application to extreme snow loads in Slovakia.

7.2.2. Conditional extremal events

Participant: Stéphane Girard.

Joint work with: L. Gardes (Univ. Strasbourg) and J. Elmethni (Univ. Paris 5)

The goal of the PhD theses of Alexandre Lekina and Jonathan El Methni was to contribute to the development of theoretical and algorithmic models to tackle conditional extreme value analysis, *ie* the situation where some covariate information X is recorded simultaneously with a quantity of interest Y. In such a case, the tail heaviness of Y depends on X, and thus the tail index as well as the extreme quantiles are also functions of the covariate. We combine nonparametric smoothing techniques [67] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles [61].

7.2.3. 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: VaR(α) 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 [61], 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 (see paragraph 7.2.2).

A second possible coherent alternative risk measure is based on expectiles [18]. 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.

7.2.4. Level sets estimation

Participant: Stéphane Girard.

Joint work with: G. Stupfler (Univ. Nottingham, UK).

The boundary bounding the set of points is viewed as the larger level set of the points distribution. This is then an extreme quantile curve estimation problem. We proposed estimators based on projection as well as on kernel regression methods applied on the extreme values set, for particular set of points [10]. We also investigate the asymptotic properties of existing estimators when used in extreme situations. For instance, we have established that the so-called geometric quantiles have very counter-intuitive properties in such situations [21] and thus should not be used to detect outliers.

7.2.5. Approximate Bayesian inference

Participant: Julyan Arbel.

Joint work with: Igor Prünster, Stefano Favaro.

Approximate Bayesian inference was tackled from two perspectives.

First, from a computational viewpoint, we have proposed an algorithm which allows for controlling the approximation error in Bayesian nonparametric posterior sampling. In [14], we show that completely random measures (CRM) represent the key building block of a wide variety of popular stochastic models and play a pivotal role in modern Bayesian Nonparametrics. The popular Ferguson & Klass representation ofCRMs as a random series with decreasing jumps can immediately be turned into an algorithm for sampling realizations of CRMs or more elaborate models involving transformedCRMs. However, concrete implementation requires to truncate the random series at some threshold resulting in an approximation error. The goal of this work is to quantify the quality of the approximation by a moment-matching criterion, which consists in evaluating a measure of discrepancy between actual moments and moments based on the simulation output. Seen as a function of the truncation level, the methodology can be used to determine the truncation level needed to reach a certain level of precision. The resulting moment-matching Ferguson & Klass algorithm is then implemented and illustrated on several popular Bayesian nonparametric models.

In [57], we focus on the truncation error of a superposed gamma process in a decreasing order representation. As in [14], we utilize the constructive representation due to Ferguson and Klass which provides the jumps of the series in decreasing order. This feature is of primary interest when it comes to sampling since it minimizes the truncation error for a fixed truncation level of the series. We quantify the quality of the approximation in two ways. First, we derive a bound in probability for the truncation error. Second, we study a moment-matching criterion which consists in evaluating a measure of discrepancy between actual moments of the CRM and moments based on the simulation output. This work focuses on a general class of CRMs, namely the superposed gamma process, which suitably transformed have already been successfully implemented in Bayesian Nonparametrics. To this end, we show that the moments of this class of processes can be obtained analytically.

Second, we have proposed an approximation of Gibbs-type random probability measures at the level of the predictive probabilities. Gibbs-type random probability measures are arguably the most “natural" generalization of the Dirichlet process. Among them the two parameter Poisson–Dirichlet process certainly stands out for the mathematical tractability and interpretability of its predictive probability, which made it the natural candidate in numerous applications, e.g., machine learning theory, probabilistic models for linguistic applications, Bayesian nonparametric statistics, excursion theory, measure-valued diffusions in population genetics, combinatorics and statistical physics. Given a sample of size n, in this work we show that the predictive probabilities of any Gibbs-type prior admit a large n approximation, with an error term vanishing as $o(1/n)$, which maintains the same desirable features as the predictive probabilities of the two parameter Poisson–Dirichlet prior.

7.2.6. Bayesian nonparametric posterior asymptotic behavior

Participant: Julyan Arbel.

Joint work with: Olivier Marchal, Stefano Favaro, Bernardo Nipoti, Yee Whye Teh.

In [24], we obtain the optimal proxy variance for the sub-Gaussianity of Beta distribution, thus proving upper bounds recently conjectured by Elder (2016). We provide different proof techniques for the symmetrical (around its mean) case and the non-symmetrical case. The technique in the latter case relies on studying the ordinary differential equation satisfied by the Beta moment-generating function known as the confluent hypergeometric function. As a consequence, we derive the optimal proxy variance for the Dirichlet distribution, which is apparently a novel result. We also provide a new proof of the optimal proxy variance for the Bernoulli distribution, and discuss in this context the proxy variance relation to log-Sobolev inequalities and transport inequalities.

The article [13] deals with a *Bayesian nonparametric inference for discovery probabilities: credible intervals and large sample asymptotics*. Given a sample of size n from a population of individual belonging to different species with unknown proportions, a popular problem of practical interest consists in making inference on the probability $D_n(l)$ that the $(n+1)$-th draw coincides with a species with frequency l in the sample, for any $l = 0, 1, ..., n$. We explore in this work a Bayesian nonparametric viewpoint for inference of $D_n(l)$. Specifically, under the general framework of Gibbs-type priors we show how to derive credible intervals for the Bayesian nonparametric estimator of $D_n(l)$, and we investigate the large n asymptotic behavior of such an estimator. We also compare this estimator to the classical Good–Turing estimator.

7.2.7. A Bayesian nonparametric approach to ecological risk assessment

Participant: Julyan Arbel.

Joint work with Guillaume Kon Kam King, Igor Prünster.

We revisit a classical method for ecological risk assessment, the Species Sensitivity Distribution (SSD) approach, in a Bayesian nonparametric framework. SSD is a mandatory diagnostic required by environmental regulatory bodies from the European Union, the United States, Australia, China etc. Yet, it is subject to much scientific criticism, notably concerning a historically debated parametric assumption for modelling species variability. Tackling the problem using nonparametric mixture models, it is possible to shed this parametric assumption and build a statistically sounder basis for SSD. We use Normalized Random Measures with Independent Increments (NRMI) as the mixing measure because they offer a greater flexibility than the Dirichlet process. Indeed, NRMI can induce a prior on the number of components in the mixture model that is less informative than the Dirichlet process. This feature is consistent with the fact that SSD practitioners do not usually have a strong prior belief on the number of components. In this work, we illustrate the advantage of the nonparametric SSD over the classical normal SSD and a kernel density estimate SSD on several real datasets[59].

7.2.8. Machine learning methods for the inversion of hyperspectral images

Participant: Stéphane Girard.

Joint work with: S. Douté (IPAG, Grenoble), M. Fauvel (INRA, Toulouse) and L. Gardes (Univ. Strasbourg).

We address the physical analysis of planetary hyperspectral images by massive inversion [58]. A direct radiative transfer model that relates a given combination of atmospheric or surface parameters to a spectrum is used to build a training set of synthetic observables. The inversion is based on the statistical estimation of the functional relationship between parameters and spectra. To deal with high dimensionality (image cubes typically present hundreds of bands), a two step method is proposed, namely K-GRSIR. It consists of a dimension reduction step followed by a regression with a non-linear least-squares algorithm. The dimension reduction is performed with the Gaussian Regularized Sliced Inverse Regression algorithm, which finds the most relevant directions in the space of synthetic spectra for the regression. The method is compared to several algorithms: a regularized version of k-nearest neighbors, partial least-squares, linear and non-linear

support vector machines. Experimental results on simulated data sets have shown that non-linear support vector machines is the most accurate method followed by K-GRSIR. However, when dealing with real data sets, K-GRSIR gives the most interpretable results and is easier to train.

7.2.9. 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.

7.2.10. Extraction and data analysis toward "industry of the future"

Participants: Florence Forbes, Hongliang Lu, Fatima Fofana, Gildas Mazo, 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.

7.3. 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.

7.3.2. Robust graph estimation

Participants: Karina Ashurbekova, Florence Forbes.

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

Graphs are an intuitive way of representing and visualizing the relationships between many variables. A graphical model is a probabilistic model whose conditional independence or other measures of relationship between random variables is given by a graph. Learning graphical models using their observed samples is an important task, and involves both structure and parameter estimation. Generally, graph estimation consists of several steps. First of all, we do not know the distribution of the real data. But we can do an assumption about this distribution. Then the measure of relationship between variables we are interested in can be chosen based on this assumption. All these measures of relationship are related with elements of the covariance or precision matrices. After estimating the covariance/precision matrix the

final graph can be constructed based on elements of this matrix. A lot of graph estimation methods rely on the Gaussian graphical model, in which the random vector Y is assumed to be Gaussian. Under this assumption, the most popular method is the graphical lasso (glasso). In practice, data may deviate from the Gaussian model in various ways. Outliers and heavy tails frequently occur. Contamination of a handful of variables in a few experiments can lead to a drastically wrong graph. So one of our objective is to deal with heavy tailed data using a new family of multivariate heavy-tailed distributions [8] and infer a graph robust to outliers without having to remove them.

7.3.3. 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 [15], 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.

7.3.4. 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 will be presented at the SPIE medical imaging 2018 conference in Houston, USA [44].

7.3.5. 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.

7.3.6. Automated ischemic stroke lesion MRI quantification

Participant: Florence Forbes.

Joint work with: Senan Doyle (Pixyl), Assia Jaillard (CHUGA) , Olivier Heck (CHUGA) , Olivier Detante (CHUGA) and Michel Dojat (GIN)

Manual delineation by an expert is currently the gold standard for lesion quantification, but is resource-intensive, suffers from inter-rater and intra-rater variability, and does not scale well to large population cohorts. We develop an automated lesion quantification method to assess the efficacy of cell therapy in patients after ischemic stroke. A high-quality sub-acute and chronic stroke dataset was supplied by HERMES. T1-w and 3D-Flair MRIs were acquired from 20 ischemic stroke patients with MCA infarct at 2 and 6 months post-event. Manual delineation was performed by an expert using the Flair image. We propose an unsupervised method employing a hidden Markov random field, with innovations to address the challenges posed by stroke MR scans. We introduce a probabilistic vascular territory atlas, adapted to the patient-specific data in a joint segmentation and registration framework, to model the potential progression and delimitation of vascular accidents. After segmentation, a good correlation is observed between manual and automated lesion volume delineation for the two time points. We therefore propose an unsupervised method with the hypothesis that such a class of methods is more robust to the diversity of images obtained with different sequence parameters and scanners; a particularly sensitive point for multi-center studies. Interestingly, this approach will be used in the European RESSTORE cohort.

7.3.7. PyHRF: A python library for the analysis of fMRI data based on local estimation of hemodynamic response function

Participants: Florence Forbes, Jaime Eduardo Arias Almeida, Aina Frau Pascual.

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

Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that allows the non-invasive study of brain function. It is based on the hemodynamic changes induced by cerebral 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 induces an increase in BOLD signal. In fact, the signal is convoluted by 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 focused on PyHRF, a software 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 constant in the entire brain, improving thus the reliability of the results. We investigated a number of real data case to demonstrate that PyHRF was a suitable tool for clinical applications. This implied the definition of guidelines to set some of the parameters required to run the software. We investigated a calibration method by comparing results with the standard SPM software in the case of a fixed HRF. An overview of the package and its performance was presented at the 16th Python in Science Conference (SciPy 2017) in Austin, TX, United States [38].

7.3.8. 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 [55]. 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 developing integrated models coupling EEG and eye movements within one single HMM for better identification of the phases. Here, the coupling should incorporate 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 must be added into the model, for the sake of synchronizing both processes.

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 develop and implement a model for jointly analyzing eye-movements and EEGs in order to improve the discrimination of the reading strategies.

7.3.9. Markov models for the analysis of the alternation of flowering in apple tree progenies

Participant: Jean-Baptiste Durand.

This research theme is supported by a Franco-German ANR grant (AlternApp project).

Joint work with: Evelyne Costes (INRA AGAP, AFEF team)

A first study was published to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoots, which correspond to portions of stems that were grown during the same year). One or several replications of each genotype were available.

Three families of indices were proposed for early detection of alternation during the juvenile phase. The first family was based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. The second family was based on a 2nd-order Markov chain with fixed and random effect in transition probabilities. The third family was based on entropy indices, in which flowering probabilities were corrected from fixed effects using Generalized Linear Models.

This allowed early quantification of alternation from the yearly numbers of inflorescences at tree scale. Some quantitative trait loci (QTL) were found in relation with these indices [40], [20].

New data sets where collected in other F1 progenies. Ancestral relationships between parents of different progenies were taken into account to enhance the power of QTL detection using Bayesian methods. Other QTLs are expected to be found using these new indices and genetic material. However, the amount of replicate per genotype and of data per replicate is quite reduced compared to those of our previous work. This is why we will investigate the loss of power in QTL detection due to a degraded amount of data, by simulating data deletion in our reference results.

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 (2016-2019). 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 was of 140 keuros

Contract with VALEO. S. Girard and A. Clément 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 was of 15 keuros.

Contract with PIXYL P. Rubini was hired for 18 months for a software valorization task regarding brain MRI segmentation. The financial support for MISTIS was of 63.5keuros

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Grenoble Idex projects

MISTIS is involved in a newly accepted transdiciplinary project **NeuroCoG**.

F. Forbes is also responsible for a work package in another project entitled **Grenoble Alpes Data Institute**.

MISTIS is also involved in a newly accepted cross-disciplinary project (CDP) RISK@UGA.

- 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 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, two 2-years multi-disciplinary projects were granted in November 2017 to Mistis in collaboration respectively with Team Necs from Inria and Gipsa-lab (DATASAFE project: understanding Data Accidents for TrAffic SAFEty) and with IPAG and Univ. Paris Sud Orsay (Regression techniques for Massive Mars hyperspectral image analysis from physical model inversion), 9 keuros each.
- Also in the context of the Grenoble Alpes Data Institute, Julyan Arbel and Stéphane Girard were awarded a funding from IRS (Initiatives de Recherche Stratégique) for a research project dedicated to extreme and Bayesian statistics, 8 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. CNRS fundings

- **Defi Mastodons, La qualité des données dans le Big Data (2015-17)**. S. Girard is involved in a 2-year project entitled “Classification de Données Hétérogènes avec valeurs manquantes appliquée au Traitement des Données Satellitaires en écologie et Cartographie du Paysage” [53], the other partners being members of Modal (Inria Lille Nord-Europe) or ENSAT-Toulouse. The total funding is 17,5 keuros.
- Stéphane Girard and Julyan Arbel were awarded a funding from TelluS-Insmi (with IPAG and Univ. Paris-Descartes), for a 1-year project entitled “unsupervised classification in high dimension", 7000 euros.
- **Defi Imag’IN MultiPlanNet (2015-2017).** This is a 2-year project to build a network for the analysis and fusion of multimodal data from planetology. There are 8 partners: IRCCYN Nantes, GIPSA-lab Grenoble, IPAG Grenoble, CEA Saclay, UPS Toulouse, LGL Lyon1, GEOPS University Orsay and Inria Mistis. F. Forbes is in charge of one work package entitled *Massive inversion of multimodal data*. Our contribution will be based on our previous work in the VAHINE project on hyperspectral images and recent developments on inverse regression methods. The CNRS support for the network is of 20 keuros. A 2-day workshop was organized in November 2017 in Grenoble, on the analysis of multimodal data for planets observation and exploration.

9.1.4. GDR Madics

Apprentissage, opTimisation à Large-échelle et cAlcul diStribué (ATLAS). Mistis is participating to this action supported by the GDR in 2016 (3 keuros).

9.1.5. Networks

MSTGA and AIGM INRA (French National Institute for Agricultural Research) networks: F. Forbes is a member 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. F. Forbes co-organized and hosted 2 of the network meetings in 2008 and 2015 in Grenoble.

9.2. International Initiatives

9.2.1. Inria Associate Teams Not Involved in an Inria International Labs

9.2.1.1. SIMERGE

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

International Partner (Institution - Laboratory - Researcher):

UGB (Senegal) - LERSTAD - Abdou Ka Diongue

Starting year: 2015

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

Entered in the LIRIMA in January 2015, this team federates researchers from LERSTAD (Laboratoire d’Etudes et de Recherches en Statistiques et Développement, Université Gaston Berger), on the one part, and MISTIS (Inria Grenoble Rhône-Alpes) on the other part. This project consolidates the existing collaborations between these two Laboratories.

The team also involves statisticians from EQUIPPE laboratory (Economie QUantitative Intégration Politiques Publiques Econométrie, Université de Lille) and associated members of Modal (Inria Lille Nord-Europe) as well as an epidemiologist from IRD (Institut de Recherche pour le Développement) at Dakar.

The following two research themes are developed : (1) Spatial extremes with application to management of extreme risks ; (2) Classification with application to global epidemiology.

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 2017 are with:

- F. Durante, Free University of Bozen-Bolzano, Italy.
- K. Qin, H. Nguyen and 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. Stehlik from Johannes Kepler Universitat Linz, Austria and Universidad de Valparaiso, Chile.
- 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, Igor Prünster, University of Turin, Italy.
- Bernardo Nipoti, Trinity College Dublin, Ireland.
- Yeh Whye Teh, Oxford University, UK.
- Stephen Walker, University of Texas at Austin, USA.

9.3. International Research Visitors

9.3.1. Visits of International Scientists

- Seydou Nourou Sylla (Université Gaston Berger, Sénégal) has been hosted by the MISTIS team for two months.
- Aboubacrène Ahmad (Université Gaston Berger, Sénégal) has been hosted by the MISTIS team for two months.
- Hien Nguyen from La Trobe university, Melbourne Australia, has been hosted for 2 days.

9.3.2. Visits to International Teams

9.3.2.1. Research Stays Abroad

- F. Forbes spent 2 weeks in April 2017 in Australia, visiting Brisbane and Melbourne universities.
- J. Arbel spent 3 months at the University of Texas at Austin.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Stéphane Girard was a member of the organization committee of the international conference "Mathematical Methods in Reliability", Grenoble, MMR2017. He also organized a session on Extremes, safety and reliability.
- Stéphane Girard and Julyan Arbel co-organized the one week 2017 school of statistics for astrophysics on Bayesian methodology, Autrans, Stat4Astro 2017.

- F. Forbes co-organized the Multiplanet 2 day workshop in November 2017 in Grenoble, on the analysis of multimodal data for planets observation and exploration.
- J.-B. Durand co-organized the CFIES conference in September 2017 in Grenoble, on teaching statistics.

10.1.2. Journal

10.1.2.1. 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.
- F. Forbes is Associate Editor of the journal Frontiers in ICT: Computer Image Analysis since its creation in Sept. 2014. Computer Image Analysis is a new specialty section in the community-run openaccess journal Frontiers in ICT. This section is led by Specialty Chief Editors Drs Christian Barillot and Patrick Bouthemy.
- In 2017, J.-B. Durand has been a guest Associate Editor of the *PLOS Computational Biology* journal.

10.1.2.2. Reviewer - Reviewing Activities

In 2017, S. Girard has been a reviewer for *Statistics and Risk Modeling*, *Extremes* and *Electronic Journal of Statistics*.

In 2017, F. Forbes has been a reviewer for *Journal of Multivariate Analysis, Computational Statistics and Data Analysis, Journal of graphical and computational statistics, Statistical analysis and data mining* .

In 2017, Julyan Arbel has been reviewer for *the Annals of Statistics, Bayesian Analysis, Biometrics, the Canadian Journal of Statistics, Statistics and Computing, Statistics and Probability Letters, the Journal of Nonparametric Statistics, the Scandinavian Journal of Statistics*, as well as for Machine Learning Conferences: *the Conference On Learning Theory (COLT), the AAAI Conference on Artificial Intelligence (AAAI), the International Conference on Learning Representations (ICLR)*. He is also writing Mathematical Reviews for MathSciNet.

10.1.3. Invited Talks

Stéphane Girard has been invited to give a talk to the following conferences:

- 10th International Conference of the ERCIM WG on Computing and Statistics [31], London, UK.

- 10th International Conference on Extreme Value Analysis [32], Delft, Netherlands.

- 27th Annual Conference of the International Environmetrics Society [33], Bergame, Italy.

- Laboratoire de Statistique Théorique et Appliquée (LSTA), Univ Paris 6. Estimation de mesures de risques à partir des Lp-quantiles extrêmes, mai 2017.

- Laboratoire des Écoulements Géophysiques et Industriels (LEGI), Univ Grenoble-Alpes, Introduction à la statistique des valeurs extrêmes, novembre 2017.

Florence Forbes has been invited to give talks at :

- University of Queensland, Brisbane, Australia, April 2017 on *Student Sliced Inverse Regression*.

- University of La Trobe, Melbourne, Australia, April 2017 on *inverse regression approach to robust nonlinear high-to-low dimensional mapping*.

- The American Statistical Association Joint Statistical meeting 2017 in Baltimore, USA - for a special session *entitled "New Dimension Reduction Methods with Applications to Biomedical Studies"* , [35].

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

- Statistics Seminar, University of Kent, Canterbury, Kent, England, November 2. Talk: Approximating predictive probabilities of Gibbs-type priors.

- Workshop 'New challenges in statistics for social sciences', Ca' Foscari University of Venice, Italy, October 16-17. Invited tutorial: Bayesian nonparametric mixture models and clustering.

- School of Statistics for Astrophysics: Bayesian methodology, Autrans, France, October 9-13. Tutorial: Bayesian nonparametric clustering.

- Journées Scientifiques d'Inria, Sophia Antipolis, France, June 14-16. Invited talk: Probabilités de découverte d'espèces: Bayes à la rescousse de Good & Turing.

- Statistics Seminar, Université du Québec à Montréal, May 25. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Statistics Seminar, Université de Sherbrooke, Canada, May 23. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Statistical Science Seminar Series, Duke University, Durham, April 14. Invited talk: Bayesian nonparametric inference for discovery probabilities.

Julyan Arbel gave also the following contributed talks:

- 10th International Conference of Computational and Methodological Statistics (ERCIM), University of London, UK, December 16-18. Invited talk: Approximating predictive probabilities of Gibbs-type priors.

- Bayes in Grenoble reading group, Grenoble, France, November 15. Talk: Approximate Bayesian computation.

- Mathematical Methods of Modern Statistics, CIRM, Luminy, France, July 10-14. Talk: Investigating predictive probabilities of Gibbs-type priors. Poster: On the sub-Gaussianity of the Beta and Dirichlet distributions.

- 11th Conference on Bayesian Nonparametrics, Paris, France, June 26-30. Poster: Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models.

- Workshop YES VIII, Eindhoven, Netherlands, January 23-25. Talk: Bayesian nonparametric inference for discovery probabilities.

Jean-Baptiste Durand has been invited to give a talk:

- at the seminar of probability and statistics at Laboratoire J. A. Dieudonné, in Nice, February 2017.

Alexis Arnaud gave a talk at:

- Congrés National d'Imagerie du Vivant, in Paris, November 2017, on *Suivi de l'hétérogénéité de la croissance de 4 modèles de gliomes par IRM multiparamétrique analysée par clustering*, [48].

Pierre-Antoine Rodesch gave a talk at:

- GDR ISIS, in Paris, March 2017, *Un algorithme one-step de reconstruction tomographique en Imagerie X spectrale*.

10.1.4. Seminars organization

- MISTIS participates in the weekly statistical seminar of Grenoble. Jean-Baptiste Durand is in charge of the organization and several lecturers have been invited in this context.
- F. Forbes and J. Arbel are co-organizing a monthly reading group on Bayesian statistics.

10.1.5. Leadership within the Scientific Community

Stéphane Girard is at the head of the associated team SIMERGE (*Statistical Inference for the Management of Extreme Risks and Global Epidemiology*) created in 2015 between MISTIS and LERSTAD (Université Gaston Berger, Saint-Louis, Sénégal). The team is part of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées), http://mistis.inrialpes.fr/simerge.

10.1.6. Scientific Expertise

Stéphane Girard was a member of the HCERES committee for the evaluation of the SAMM laboratory, Université Paris 1. He also was a referee for the NWO, Netherlands Organisation for Scientific Research.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : Stéphane Girard, *Statistique Inférentielle Avancée*, 18 ETD, M1 level, Ensimag. Grenoble-INP, France.

Master: Jean-Baptiste Durand, *Statistics and probability*, 192 ETD, M1 and M2 levels, Ensimag Grenoble INP, France. Head of the MSIAM M2 program, in charge of the statistics and data science tracks ([64]).

J.-B. Durand is a faculty member at Ensimag, Grenoble-INP.

Master and PhD course: Julyan Arbel gave a course on *Bayesian nonparametric statistics*, 25 ETD, Inria Montbonnot, France.

J-M. Bécu, C. Albert are teaching at UGA.

Licence : Alexis Arnaud, *Modélisations mathématiques*, 48 ETD, L2 level, IUT2 Grenoble. Université Grenoble Alpes, France.

Licence : Alexis Arnaud, *Analyse pour l'ingénieur*, 33 ETD, L3 level, Ensimag. Grenoble-INP, France.

Licence : Alexis Arnaud, *Soutien en Analyse pour l'ingénieur*, 39 ETD, L3 level, Ensimag. Grenoble-INP, France.

Licence: Brice Olivier, *Probabilités pour l'informatique*, 27 ETD, M1 level, Ensimag. Grenoble-INP, France.

Licence: Brice Olivier, *Principes et méthodes statistiques*, 36 ETD, L3 level, Ensimag. Grenoble-INP, France.

Licencel: Brice Olivier, *Retours d'expériences (ReX)*, 2 ETD, M2 level, Ensimag. Grenoble-INP, France.

10.2.2. Supervision

PhD: Maïlys Lopes, "*Suivi écologique des prairies semi-naturelles : analyse statistique de séries temporelles denses d'images satellite à haute résolution spatiale*", defended November 2017, Stéphane Girard and Mathieu Fauvel (INRA Toulouse).

PhD in progress: "*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: Thibaud Rahier, "*Data-mining pour la fusion de données structurées et non-structurées*", started on November 2015, Florence Forbes and Stéphane Girard.

PhD in progress: Clément Albert, "*Limites de crédibilité d'extrapolation des lois de valeurs extrêmes*", started on January 2016, Stéphane Girard.

PhD in progress: Alexis Arnaud "*Multiparametric MRI statistical analysis for the identification and follow-up of brain tumors* ", October 2014, Florence Forbes, Benjamin Lemasson and Emmanuel Barbier (GIN).

PhD in progress: Pierre-Antoine Rodesch, "*Spectral tomography and tomographic reconstruction algorithms*", October 2015, Florence Forbes, Clarisse Fournier and Veronique Rebuffel (CEA Leti Grenoble).

PhD in progress: Brice Olivier, "*Joint analysis of eye-movements and EEGs using coupled hidden Markov and topic models*", October 2015, Jean-Baptiste Durand, Marianne Clausel and Anne Guérin-Dugué (Université Grenoble Alpes).

PhD in progress: Karina Ashurbekova, "*Robust Graphical models*", October 2016, Florence Forbes and Sophie Achard (Gipsa-lab, Grenoble).

PhD in progress: Veronica Munoz Ramirez, "*Extraction de signatures dans les données IRM de patients parkinsonioniens de novo*", October 2017, Florence Forbes, Julyan Arbel and Michel Dojat (GIN).

PhD in progress: Fabien Boux, "*Développement de méthodes statistiques pour l'imagerie IRM fingerprinting* ", September 2017, Florence Forbes, Julyan Arbel and Emmanuel Barbier (GIN).

10.2.3. Juries

- S. Girard was a member of 3 PhD committees in 2017:

  - Mohamed Néjib Dalhoumi, *Sur l'estimation de probabilités de queues multivariées*, Univ. Montpellier, September 2017.

  - Achmad Choiruddin, *Sélection de variables pour des processus ponctuels spatiaux*, Univ. Grenoble, September 2017.

  - Patricia Tencaliec, *Development in statistics applied to hydrometeorology: imputation of streamflow data and semiparametic precipitation modeling*, Univ. Grenoble, February 2017.

- Florence Forbes has been reviewer of 2 PhD thesis in 2017:

  - Julie Aubert, *Analyse statistique de données biologiques à haut débit*, AgroParisTech, January 2017.

  - Adrien Faivre, *Analyse d'images hyperspectrales*, University of Besançon Franche-Comté, December 14, 2017.

- F. Forbes was a member of 3 PhD committees in 2017:

  - Clément Elvira, *Modèles bayesiens pour l'identification de représentations antiparcimonieuses et l'analyse en composantes principales bayésienne non paramétrique*, November 10, 2017, Centrale Lille

  - Melanie Bernard, *Système modulaire de traitement pour la tomographie d'émission à partir de détecteurs CdZnTe*, November 6, 2017, CEA-Leti, Grenoble.

  - Vincent Drouard, *, Localisation et suivi de visages à partir d'images et de sons*, December 18, 2017, Inria Grenoble.

- Julyan Arbel acted as a reviewer for the PhD thesis of Ilaria Bianchini, *Modeling and computational aspects of dependent completely random measures in Bayesian nonparametric statistics*, December 2017, Politecnico di Milano, Italy.

10.2.3.1. Other committees

- Grenoble Pole Cognition. F. Forbes is representing Inria and LJK in the pole.
- PRIMES Labex, Lyon. F. Forbes is a member of the strategic committee. F. Forbes is representing Inria since 2016.
- F. Forbes is a member of the executive committee of the Idex CDP Grenoble Data institute.
- F. Forbes is a member of the Committee for technological project and engineer candidate selection at Inria Grenoble Rhône-Alpes ("Commission du développement technologique ") since 2015.
- F. Forbes is a member of the " Comité d'Organisation Stratégique" (COS) since September 2017.
- F. Forbes has been a member of 3 selection committees, 2 for Professors at Centrale Lille and at Paris-Descartes 5, and 1 for assistant professors at University of Lille.
- F. Forbes has been a member of the committee awarding *Grand prix Inria de l'académie des sciences*, June 2017.
- 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 is a member of the "Comité des Emplois Scientifiques" at Inria Grenoble Rhône-Alpes since 2015.

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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 2, pp. 369–384

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. LOPES. *Ecological monitoring of semi-natural grasslands: statistical analysis of dense satellite image time series with high spatial resolution*, Institut National Polytechnique de Toulouse, November 2017, https://hal.inria.fr/tel-01684474

Articles in International Peer-Reviewed Journals

[12] M. ALBUGHDADI, L. CHAARI, J.-Y. TOURNERET, F. FORBES, P. CIUCIU. *A Bayesian Non-Parametric Hidden Markov Random Model for Hemodynamic Brain Parcellation*, in "Signal Processing", June 2017, vol. 135, pp. 132–146 [*DOI :* 10.1016/J.SIGPRO.2017.01.005], https://hal.archives-ouvertes.fr/hal-01426385

[13] J. ARBEL, S. FAVARO, B. NIPOTI, Y. W. TEH. *Bayesian nonparametric inference for discovery probabilities: credible intervals and large sample asymptotics*, in "Statistica Sinica", January 2017, vol. 27, pp. 839-858, https://arxiv.org/abs/1506.04915 [*DOI :* 10.5705/SS.202015.0250], https://hal.archives-ouvertes.fr/hal-01203324

[14] J. ARBEL, I. PRÜNSTER. *A moment-matching Ferguson & Klass algorithm*, in "Statistics and Computing", June 2017, vol. 27, n° 1, pp. 3-17, https://arxiv.org/abs/1606.02566 [*DOI :* 10.1007/s11222-016-9676-8], https://hal.archives-ouvertes.fr/hal-01396587

[15] 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", 2018, forthcoming, https://hal.archives-ouvertes.fr/hal-01545548

[16] A. CHIANCONE, F. FORBES, S. GIRARD. *Student Sliced Inverse Regression*, in "Computational Statistics and Data Analysis", September 2017, vol. 113, pp. 441-456 [*DOI :* 10.1016/J.CSDA.2016.08.004], https://hal.archives-ouvertes.fr/hal-01294982

[17] A. CHIANCONE, S. GIRARD, J. CHANUSSOT. *Collaborative Sliced Inverse Regression*, in "Communication in Statistics - Theory and Methods", 2017, vol. 46, n° 12, pp. 6035–6053 [*DOI :* 10.1080/03610926.2015.1116578], https://hal.inria.fr/hal-01158061

[18] A. DAOUIA, S. GIRARD, G. STUPFLER. *Estimation of Tail Risk based on Extreme Expectiles*, in "Journal of the Royal Statistical Society: Series B", 2017 [*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", 2017, forthcoming, https://hal.inria.fr/hal-01585215

[20] J.-B. DURAND, A. ALLARD, B. GUITTON, E. VAN DE WEG, M. BINK, E. COSTES. *Predicting Flowering Behavior and Exploring Its Genetic Determinism in an Apple Multi-family Population Based on Statistical Indices and Simplified Phenotyping*, in "Frontiers in Plant Science", June 2017, vol. 8, pp. 858–872 [*DOI :* 10.3389/FPLS.2017.00858], https://hal.inria.fr/hal-01564977

[21] S. GIRARD, G. STUPFLER. *Intriguing properties of extreme geometric quantiles*, in "REVSTAT - Statistical Journal", January 2017, vol. 15, n° 1, pp. 107–139, https://hal.inria.fr/hal-00865767

[22] M. LOPES, M. M. FAUVEL, S. GIRARD, D. SHEEREN. *Object-based classification of grasslands from high resolution satellite image time series using Gaussian mean map kernels*, in "Remote Sensing", July 2017, vol. 9, n° 7, Article 688 [*DOI :* 10.3390/RS9070688], https://hal.inria.fr/hal-01424929

[23] M. LOPES, M. FAUVEL, A. OUIN, S. GIRARD. *Spectro-Temporal Heterogeneity Measures from Dense High Spatial Resolution Satellite Image Time Series: Application to Grassland Species Diversity Estimation*, in

"Remote Sensing", October 2017, vol. 9, n^{o} 10, pp. 993:1-23 [*DOI :* 10.3390/RS9100993], https://hal.archives-ouvertes.fr/hal-01613722

[24] O. MARCHAL, J. ARBEL. *On the sub-Gaussianity of the Beta and Dirichlet distributions*, in "Electronic Communications in Probability", October 2017, vol. 22, pp. 1-14, https://arxiv.org/abs/1705.00048 , https://hal.archives-ouvertes.fr/hal-01521300

[25] 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-01652011

[26] M. STEHLIK, P. AGUIRRE, S. GIRARD, P. JORDANOVA, J. KISEĽÁK, S. TORRES-LEIVA, Z. SADOVSKY, A. RIVERA. *On ecosystems dynamics*, in "Ecological Complexity", March 2017, vol. 29, pp. 10–29 [*DOI :* 10.1016/J.ECOCOM.2016.11.002], https://hal.inria.fr/hal-01394734

Invited Conferences

[27] J. ARBEL. *Approximating predictive probabilities of Gibbs-type priors*, in "ERCIM - 10th International Conference of the ERCIM WG on Computational and Methodological Statistics", London, United Kingdom, December 2017, https://hal.archives-ouvertes.fr/hal-01667746

[28] J. ARBEL. *Bayesian nonparametric clustering*, in "School of Statistics for Astrophysics: Bayesian methodology", Autrans, France, October 2017, https://hal.archives-ouvertes.fr/hal-01667760

[29] J. ARBEL. *Bayesian nonparametric mixture models and clustering*, in "Workshop 'New challenges in statistics for social sciences'", Venise, Italy, October 2017, https://hal.archives-ouvertes.fr/hal-01667755

[30] J. ARBEL. *Probabilités de découverte d'espèces: Bayes à la rescousse de Good & Turing*, in "Journées Scientifiques d'Inria", Sophia Antipolis, France, June 2017, https://hal.archives-ouvertes.fr/hal-01667788

[31] S. GIRARD, A. DAOUIA, G. STUPFLER. *Extreme M-quantiles as risk measures*, in "10th International Conference of the ERCIM WG on Computing and Statistics", London, United Kingdom, December 2017, https://hal.archives-ouvertes.fr/hal-01667201

[32] S. GIRARD, L. GARDES. *Estimation of the functional Weibull tail-coefficient*, in "10th International Conference on Extreme Value Analysis", Delft, Netherlands, June 2017, https://hal.archives-ouvertes.fr/hal-01571990

[33] S. GIRARD, M. LOPES, M. M. FAUVEL, D. SHEEREN. *Object-based Classification of Grassland Management Practices From High Resolution Satellite Image Time Series With Gaussian Mean Map Kernels*, in "27th Annual Conference of the International Environmetrics Society", Bergame, Italy, July 2017, https://hal.archives-ouvertes.fr/hal-01571079

[34] S. GIRARD, G. STUPFLER. *Some negative results on extreme multivariate quantiles defined through convex optimisation*, in "10th International Conference of the ERCIM WG on Computing and Statistics", London, United Kingdom, December 2017, https://hal.archives-ouvertes.fr/hal-01667186

[35] C.-C. TU, F. FORBES, N. WANG, B. LEMASSON. *Structured Mixture of linear mappings in high dimension*, in "JSM 2017 - Joint Statistical Meeting", Baltimore, United States, July 2017, https://hal.inria.fr/hal-01653601

International Conferences with Proceedings

[36] C. ALBERT, A. DUTFOY, S. GIRARD. *On the extrapolation limits of extreme-value theory for risk management*, in "MMR 2017 - 10th International Conference on Mathematical Methods in Reliability", Grenoble, France, July 2017, 5 p. , https://hal.archives-ouvertes.fr/hal-01571099

[37] C. ALBERT, A. DUTFOY, S. GIRARD. *On the relative approximation error of extreme quantiles by the block maxima method*, in "10th International Conference on Extreme Value Analysis", Delft, Netherlands, June 2017, https://hal.archives-ouvertes.fr/hal-01571047

[38] J. ARIAS, P. CIUCIU, M. DOJAT, F. FORBES, A. FRAU-PASCUAL, T. PERRET, J. M. WARNKING. *PyHRF: A Python Library for the Analysis of fMRI Data Based on Local Estimation of the Hemodynamic Response Function*, in "16th Python in Science Conference (SciPy 2017)", Austin, TX, United States, July 2017 [*DOI :* 10.25080/SHINMA-7F4C6E7-006], https://hal.archives-ouvertes.fr/hal-01566457

[39] M. CANO, J. ARIAS, J. A. PÉREZ. *Session-Based Concurrency, Reactively*, in "37th International Conference on Formal Techniques for Distributed Objects, Components, and Systems (FORTE)", Neuchâtel, Switzerland, A. BOUAJJANI, A. SILVA (editors), Formal Techniques for Distributed Objects, Components, and Systems, Springer, June 2017, vol. LNCS-10321, pp. 74-91 [*DOI :* 10.1007/978-3-319-60225-7_6], https://hal.archives-ouvertes.fr/hal-01566466

[40] J.-B. DURAND, A. ALLARD, B. GUITTON, E. VAN DE WEG, M. C. A. M. BINK, E. COSTES. *Genetic determinism of flowering regularity over years in an apple multi-family population*, in "International Symposium on Flowering, Fruit Set and Alternate Bearing", Palermo, Italy, June 2017, https://hal.inria.fr/hal-01565681

[41] J.-B. DURAND. *Challenges d'analyse de données : une formation par la pratique transversale et multidisciplinaire en science des données*, in "CFIES2017 - Colloque Francophone International sur l'Enseignement de la Statistique", Grenoble, France, September 2017, https://hal.inria.fr/hal-01611032

[42] B. LEMASSON, N. COLLOMB, A. ARNAUD, E. LUC BARBIER, F. FORBES. *Monitoring glioma heterogeneity during tumor growth using clustering analysis of multiparametric MRI data*, in "ISMRM International Society for Magnetic Resonance in Medicine", Honolulu, United States, April 2017, https://hal.inria.fr/hal-01652033

[43] M. LOPES, M. FAUVEL, A. OUIN, S. GIRARD. *Potential of Sentinel-2 and SPOT5 (Take5) time series for the estimation of grasslands biodiversity indices*, in "MultiTemp 2017 - 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images", Bruges, Belgium, June 2017, pp. 1-4 [*DOI :* 10.1109/MULTITEMP.2017.8035206], https://hal.archives-ouvertes.fr/hal-01556786

[44] 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, https://hal.inria.fr/hal-01652017

[45] S. N. SYLLA, S. GIRARD, A. K. DIONGUE, A. DIALLO, C. SOKHNA. *Hierarchical kernel applied to mixture model for the classification of binary predictors*, in "61st ISI World Statistics Congress", Marrakech, Morocco, July 2017, https://hal.archives-ouvertes.fr/hal-01587163

[46] V. WATSON, J.-F. TROUILHET, F. PALETOU, S. GIRARD. *Inference of an explanatory variable from observations in a high-dimensional space: Application to high-resolution spectra of stars*, in "IEEE International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics (ECMSM)", San Sebastian, Spain, May 2017, https://hal.archives-ouvertes.fr/hal-01533227

National Conferences with Proceedings

[47] C. ALBERT, A. DUTFOY, S. GIRARD. *Etude de l'erreur relative d'approximation des quantiles extrêmes*, in "49èmes Journées de Statistique organisées par la Société Française de Statistique", Avignon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01533220

[48] F. ANDRIATSITOAINA, N. COLLOMB, A. ARNAUD, F. FORBES, J.-P. ISSARTEL, C. LOUSSOUARN, E. GARCION, E. LUC BARBIER, B. LEMASSON. *Suivi de l'hétérogénéité de la croissance de 4 modèles de gliomes par IRM multiparamétrique analysée par clustering*, in "congrès national de l'imagerie du vivant", Paris, France, November 2017, https://hal.inria.fr/hal-01652029

[49] J. ARBEL, D. FRAIX-BURNET, S. GIRARD. *Les écoles d'astrostatistique " Statistics for Astrophysics "*, in "CFIES 2017 - 5ème Colloque Francophone International sur l'Enseignement de la Statistique", Grenoble, France, September 2017, https://hal.inria.fr/hal-01583854

[50] B. LEMASSON, N. COLLOMB, A. ARNAUD, F. FORBES, E. L. BARBIER. *Suivi de l'hétérogénéité de la croissance des gliomes par IRM multiparamétrique analysée par clustering*, in "SFRMBM Societe Francaise de Resonance Magnetique en Biologie et Medecine", Bordeaux, France, March 2017, https://hal.inria.fr/hal-01652300

Conferences without Proceedings

[51] J. ARBEL. *Bayesian nonparametric inference for discovery probabilities*, in "YES VIII Workshop on Uncertainty Quantification", Eindhoven, Netherlands, January 2017, https://hal.archives-ouvertes.fr/hal-01667794

[52] J. ARBEL. *Investigating predictive probabilities of Gibbs-type priors*, in "Mathematical Methods of Modern Statistics", Marseille, France, July 2017, https://hal.archives-ouvertes.fr/hal-01667765

[53] S. IOVLEFF, M. FAUVEL, S. GIRARD, C. PREDA, V. VANDEWALLE. *Mixture Models with Missing data Classication of Satellite Image Time Series: QUALIMADOS: Atelier Qualité des masses de données scientiques*, in "Journées Science des Données MaDICS 2017", Marseille, France, June 2017, pp. 1-60, https://hal.archives-ouvertes.fr/hal-01649206

[54] M. LOPES, M. FAUVEL, A. OUIN, S. GIRARD. *Evaluation de la biodiversité des prairies semi-naturelles par télédétection hyperspectrale*, in "SFPT-GH 2017 - 5ème colloque scientifique du groupe thématique hyperspectral de la Société Française de Photogrammétrie et Télédétection", Brest, France, May 2017, vol. 24, https://hal.archives-ouvertes.fr/hal-01542063

[55] 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 "European Conference on Eye Movements - ECEM 2017", Wuppertal, Germany, August 2017, https://hal.inria.fr/hal-01671224

[56] G. STUPFLER, S. GIRARD, A. GUILLOU. *Estimating a frontier function using a high-order moments method*, in "31st European Meeting of Statisticians", Helsinki, Finland, July 2017, https://hal.archives-ouvertes.fr/hal-01571126

Scientific Books (or Scientific Book chapters)

[57] J. ARBEL, I. PRÜNSTER. *Truncation error of a superposed gamma process in a decreasing order representation*, in "Bayesian Statistics in Action", R. ARGIENTO, E. LANZARONE, I. ANTONIANO VILLALOBOS, A. MATTEI (editors), Bayesian Statistics in Action, January 2017, vol. 194, pp. 11–19, https://hal.archives-ouvertes.fr/hal-01405580

[58] M. FAUVEL, S. GIRARD, S. DOUTÉ, L. GARDES. *Machine Learning Methods for the Inversion of Hyperspectral Images*, in "Horizons in World Physics", A. REIMER (editor), Nova Science, 2017, vol. 290, pp. 51-77, https://hal.inria.fr/hal-01445638

[59] G. KON KAM KING, J. ARBEL, I. PRÜNSTER. *A Bayesian nonparametric approach to ecological risk assessment*, in "Bayesian Statistics in Action", R. ARGIENTO, E. LANZARONE, I. ANTONIANO VILLALOBOS, A. MATTEI (editors), Bayesian Statistics in Action, January 2017, vol. 194, pp. 151–159, https://hal.archives-ouvertes.fr/hal-01405593

Other Publications

[60] J. ARBEL, J.-B. SALOMOND. *Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models*, June 2017, 1 p. , BNP 2017 - 11th Conference on Bayesian NonParametrics, Poster, https://hal.archives-ouvertes.fr/hal-01667781

[61] J. EL METHNI, L. GARDES, S. GIRARD. *Kernel estimation of extreme regression risk measures*, November 2017, working paper or preprint, https://hal.inria.fr/hal-01393519

[62] D. FRAIX-BURNET, C. BOUVEYRON, S. GIRARD, J. ARBEL. *Unsupervised classification in high dimension*, June 2017, European Week of Astronomy and Space Science (EWASS 2017), Poster, https://hal.archives-ouvertes.fr/hal-01569733

[63] B. LEMASSON, N. COLLOMB, A. ARNAUD, F. FORBES, E. L. BARBIER. *Monitoring brain tumor evolution using multiparametric MRI*, April 2017, 2017 IEEE International Symposium on Biomedical Imaging, Poster, https://hal.inria.fr/hal-01652026

References in notes

[64] M.-R. AMINI, J.-B. DURAND, O. GAUDOIN, E. GAUSSIER, A. IOUDITSKI. *Data Science: an international training program at master level*, in "Statistique et Enseignement (ISSN 2108-6745)", June 2016, vol. 7, n^{o} 1, pp. 95-102, https://hal.inria.fr/hal-01342469

[65] 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

[66] P. EMBRECHTS, C. KLÜPPELBERG, T. MIKOSH. *Modelling Extremal Events*, Applications of Mathematics, Springer-Verlag, 1997, vol. 33

[67] F. FERRATY, P. VIEU. *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, 2006

[68] 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

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

[70] G. MAZO. *A semiparametric and location-shift copula-based mixture model*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01263382