IN PARTNERSHIP WITH:
Institut polytechnique de Grenoble

Université Grenoble Alpes

Activity Report 2016

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:

3.1.1. - Modeling, representation
3.1.4. - Uncertain data
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.5. - Bayesian methods
3.4.7. - Kernel methods
6.1. - Mathematical Modeling
8.2. - Machine learning
8.3. - Signal analysis

Other Research Topics and Application Domains:

1.3.1. - Understanding and simulation of the brain and the nervous system
2.6.1. - Brain imaging
3.4.1. - Natural risks
3.4.2. - Industrial risks and waste
9.9.1. - Environmental risks

1. Members

Research Scientists

Florence Forbes [Team leader, Inria, Senior researcher, HDR]
Julyan Arbel [Inria, Researcher, from Sep 2016]
Stephane Girard [Inria, Senior researcher, HDR]
Gildas Mazo [Inria, Starting Research Position, from Oct 2016]

Faculty Member

Jean-Baptiste Durand [INP Grenoble, Associate professor]

Engineers

Jaime Eduardo Arias Almeida [Inria, since Oct 2016]
Thomas Perret [Inria, until Aug 2016]
Pascal Rubini [Inria]

PhD Students

Clement Albert [Inria]
Alexis Arnaud [Univ. Grenoble I]
Alessandro Chiancone [Univ. Grenoble I, until Oct 2016]
Aina Frau Pascual [Inria]
Brice Olivier [Univ. Grenoble I]
Thibaud Rahier [Schneider Electric]
Pierre-Antoine Rodesch [CEA]
Karina Ashurbekova [Gipsa-Lab, from Oct 2016]

Seydou Nourou Sylla [Inria, Univ. Gaston Berger, St Louis, Senegal, until Jul 2016]

Post-Doctoral Fellows

Pablo Mesejo Santiago [Inria, until Aug 2016]
Jean-Michel Becu [Inria, from Apr 2016]
Emeline Perthame [Inria]

Visiting Scientist

Maïlys Lopes [INRA, Nov 2016]

Administrative Assistants

Myriam Etienne [Inria, until Sep 2016]
Marion Ponsot [Inria, from Sep 2016]

Others

Karina Ashurbekova [Gipsa-Lab, Interns, from Feb 2016 until Jul 2016]
Jules Corset [Inria, Interns, Jan 2016]
Benjamin Lemasson [Inserm, GIN, Researcher, external collaborator from Jun 2016]

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, Alessandro Chiancone, Stephane Girard, Julyan Arbel, Gildas Mazo, Jean-Michel Becu.

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: Brice Olivier, Thibaud Rahier, Jean-Baptiste Durand, Florence Forbes, Karina Ashurbekova.

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: Clement Albert, Alessandro Chiancone, Stephane Girard, Seydou Nourou Sylla, Pablo Mesejo Santiago, Florence Forbes, Emeline Perthame, Jean-Michel Becu.

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

We also consider methods which do not assume a parametric model. The approaches are non-parametric in the sense that they do not require the assumption of a prior model on the unknown quantities. This property is important since, for image applications for instance, it is very difficult to introduce sufficiently general parametric models because of the wide variety of image contents. Projection methods are then a way to decompose the unknown quantity on a set of functions (*e.g.* wavelets). Kernel methods which rely on smoothing the data using a set of kernels (usually probability distributions) are other examples. Relationships exist between these methods and learning techniques using Support Vector Machine (SVM) as this appears in the context of *level-sets estimation* (see section 3.3.2). Such non-parametric methods have become the cornerstone when dealing with functional data [77]. This is the case, for instance, when observations are curves. They enable us to model the data without a discretization step. More generally, these techniques are of great use for *dimension reduction* purposes (section 3.3.3). They enable reduction of the dimension of the functional or multivariate data without assumptions on the observations distribution. Semi-parametric methods refer to methods that include both parametric and non-parametric aspects. Examples include the Sliced Inverse Regression (SIR) method [80] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [76], which is based on the modelling of distribution tails (see section 3.3.1). It differs from traditional statistics which focuses on the central part of distributions, *i.e.* on the most probable events. Extreme value theory shows that distribution tails can be modelled by both a functional part and a real parameter, the extreme value index.

3.3.1. Modelling extremal events

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

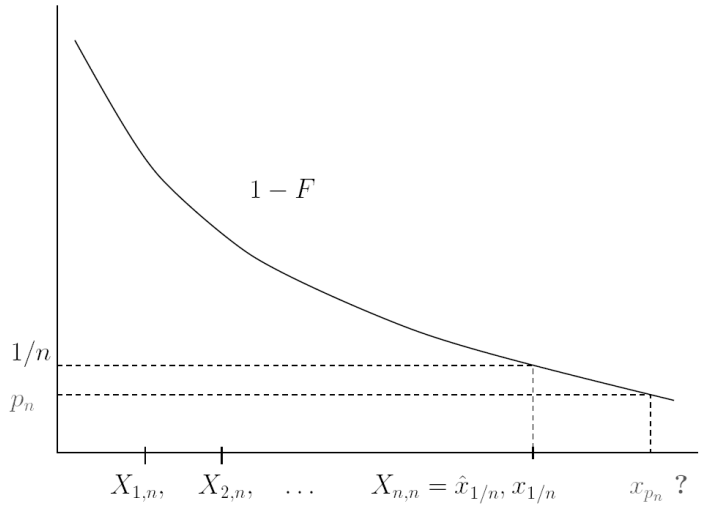

Figure 1. The curve represents the survival function $x \to P(X > x)$. The $1/n$-quantile is estimated by the maximum observation so that $\widehat{x}_{1/n} = X_{n,n}$. As illustrated in the figure, to estimate p_n-quantiles with $p_n < 1/n$, it is necessary to extrapolate beyond the maximum observation.

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 [78]. Another field of investigation is to take into account the application goal in the dimension reduction step. One of our approaches is therefore to develop new Gaussian models of high dimensional data for parametric inference [74]. 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 [80].

4. Application Domains

4.1. Image Analysis

Participants: Alexis Arnaud, Aina Frau Pascual, Florence Forbes, Stephane Girard, Pascal Rubini, Alessandro Chiancone, Thomas Perret, Pablo Mesejo Santiago, 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 Becu, Florence Forbes.

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: Pablo Mesejo Santiago, Aina Frau Pascual, Florence Forbes, Stephane Girard, Seydou Nourou Sylla, Emeline Perthame, Jean-Baptiste Durand, Clement Albert, Julyan Arbel, Jean-Michel Becu, Thibaud Rahier, Brice Olivier, Karina Ashurbekova.

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

- The Pixyl startup (http://pixyl.io) created in March 2015 by F. Forbes (Mistis) with M. Dojat (INSERM), a former Mistis post-doctoral fellow S. Doyle (CEO) and IT Translation is one of the two Inria start-ups winners of the NETVA 2016 competition. S. Doyle travelled to Washington to take part in a personalized support program to learn about the North American markets. The NETVA competition is open to French hi-tech start-ups. It is organized by the science and technology departments of the French embassies in Canada and the USA. Pixyl develops neuro-imaging software which automatically analyses brain lesion load using MRI scans, for improved decision-making during clinical trials and routine clinical use.
- Vision 4.0 FUI Minalogic project: Mistis is one of the 4 partners in the Vision 4.0 project that started in October 2016. This is one of the 8 projects funded by the Minalogic Pôle de competitivité in 2016. The support is of 3.4 Meuros.

5.1.1. Awards

- 2016 Award for Outstanding Contributions in Neural Systems. Antoine Deleforge (now with the PANAMA team, Inria Bretagne-Atlantique), Florence Forbes (MISTIS team) and Radu Horaud (PERCEPTION team) received the 2016 Hojjat Adeli Award for Outstanding Contributions in Neural Systems for their paper: A. Deleforge, F. Forbes, and R. Horaud (2015), Acoustic Space Learning for Sound-source Separation and Localization on Binaural Manifolds, International Journal of Neural Systems, 25:1,(21 pages) [75]. The Award for Outstanding Contributions in Neural Systems established by World Scientific Publishing Co. in 2010, is awarded annually to the most innovative paper published in the previous volume/year of the International Journal of Neural Systems.
- MITACS Globalink Research Award - Inria - for research in Canada. Alexis Arnaud received the MITACS award and a 5 kdollars grant to spend 5 months in the Mathematics and statistics department of McGill University in Montreal, Canada, working with Prof. Russel Steele.

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: Pablo Mesejo Santiago, Jan M Warnking 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: Health - Statistics - Brain MRI - Medical imaging - Robust clustering
FUNCTIONAL DESCRIPTION

The package implements mixtures of so-called multiple scaled Student distributions, which are generalization 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, Florence Forbes and Darren Wraith
- Contact: Florence Forbes
- URL: http://mistis.inrialpes.fr/realisations.html

6.3. PyHRF

KEYWORDS: FMRI - Statistic analysis - Neurosciences - IRM - Brain - Health - 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: Thomas Vincent, Solveig Badillo, Lotfi Chaari, Christine Bakhous, Florence Forbes, Philippe Ciuciu, Laurent Risser, Thomas Perret, Aina Frau Pascual and Jaime Eduardo Arias Almeida
- 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.

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 [71]) based on a mixture of Generalized Student distributions.

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

7. New Results

7.1. Mixture models

7.1.1. High dimensional Kullback-Leilbler divergence for supervised clustering

Participant: Stephane Girard.

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

In the PhD work of Charles Bouveyron [74], 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 [50], 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. A parsimonious model is discussed to deal with the high dimension of the data and the small sample size. A high dimensional symmetrized Kullback-Leibler divergence (KLD) is introduced to compute the similarity between each pair of grasslands. The model is positively compared to the conventional KLD to construct a positive definite kernel used in SVM for supervised classification.

7.1.2. Single-run model selection in mixtures

Participants: Florence Forbes, Alexis Arnaud.

Joint work with: Russel Steele, McGill University, Montreal, Canada.

A number of criteria exist to select the number of components in a mixture automatically based on penalized likelihood criteria (eg. AIC, BIC, ICL etc.) but they usually require to run several models for different number of components to choose the best one. In this work, the goal was to investigate existing alternatives that can select the component number from a single run and to develop such a procedure for our MRI analysis. These objectives were achieved for the main part as 1) different single run methods have been implemented and tested for Gaussian and Standard mixture models, 2) a Bayesian version of Generalized Student mixtures have been designed that allows the use of the methods in 1), and 3) we also proposed a new heuristic based on this Bayesian model that shows good performance and lower computational times. A more complete validation on simulated data and tests on real MRI data need still to be performed. The single run methods studied are based on a fully Bayesian approach involving therefore specification of appropriate priors and choice of hyperparameters. To estimate our Bayesian mixture model, we use a Variational Expectation-Maximization algorithm (VEM). For the heuristic, we add an additional step inside VEM in order to compute in parallel the corresponding VEM step with one less component. If the lower-bound of the model likelihood is higher with one less component, then we delete this component and go to the next VEM step, until convergence of the algorithm. As regards software development, the Rcpp package has been used to bridge pure R code with more efficient C++ code. This project has been initiated with Alexis Arnaud's visit to McGill University in Montreal in the context of his Mitacs award.

7.1.3. Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models

Participant: Julyan Arbel.

Joint work with: Jean-Bernard Salomond (Université Paris-Est).

In mixture models, latent variables known as allocation variables play an essential role by indicating, at each iteration, to which component of the mixture observations are linked. In sequential algorithms, these latent variables take on the interpretation of particles. We investigate the use of quasi Monte Carlo within sequential Monte Carlo methods (a technique known as sequential quasi Monte Carlo) in nonparametric mixtures for density estimation. We compare them to sequential and non sequential Monte Carlo algorithms. We highlight a critical distinction of the allocation variables exploration of the latent space under each of the three sampling approaches. This work has been presented at the *Practical Bayesian Nonparametrics* NIPS workshop [48].

7.1.4. Truncation error of a superposed gamma process in a decreasing order representation

Participant: Julyan Arbel.

Joint work with: Igor Prünster (University Bocconi, Milan).

Completely random measures (CRM) represent a key ingredient of a wealth of stochastic models, in particular in Bayesian Nonparametrics for defining prior distributions. CRMs can be represented as infinite random series of weighted point masses. A constructive representation due to Ferguson and Klass 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. This work has been presented at the *Advances in Approximate Bayesian Inference* NIPS workshop [47].

7.1.5. Non linear mapping by mixture of regressions with structured covariance matrix

Participant: Emeline Perthame.

Joint work with: Emilie Devijver (KU Leuven, Belgium) and Mélina Gallopin (Université Paris Sud).

In genomics, the relation between phenotypical responses and genes are complex and potentially non linear. Therefore, it could be interesting to provide biologists with statistical models that mimic and approximate these relations. In this paper, we focus on a dataset that relates genes expression to the sensitivity to alcohol of drosophila. In this framework of non linear regression, GLLiM (Gaussian Locally Linear Mapping) is an efficient tool to handle non linear mappings in high dimension. Indeed, this model based on a joint modeling of both responses and covariates by Gaussian mixture of regressions has demonstrated its performance in non linear prediction for multivariate responses when the number of covariates is large. This model also allows the addition of latent factors which have led to interesting interpretation of the latent factors in image analysis. Nevertheless, in genomics, biologists are more interested in graphical models, representing gene regulatory networks. For this reason, we developed an extension of GLLiM in which covariance matrices modeling the dependence structure of genes in each clusters are blocks diagonal, using tools derived for graphical models. This extension provides a new class of interpretable models that are suitable to genomics application fields while keeping interesting prediction properties.

7.1.6. Extended GLLiM model for a subclustering effect: Mixture of Gaussian Locally Linear Mapping (MoGLLiM)

Participant: Florence Forbes.

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

The work of Chun-Chen Tu and Naisyin Wang pointed out a problem with the original GLLiM model that they propose to solve with a divide-remerge method. The proposal seems to be efficient on test data but the resulting procedure does not anymore correspond to the optimization of a single statistical model. The idea of this work is then to discuss the possibility to change the original GLLiM model in order to account for sub-clusters directly. A small change in the definition seems to have such an effect while remaining tractable. However, we will probably have to be careful with potential non-identifiability issue when dealing with clusters and sub-clusters.

7.2. Semi and non-parametric methods

7.2.1. Robust estimation for extremes

Participants: Clement Albert, Stephane 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 [46]. To reduce this sensitivity, robust methods are needed and, in [21], 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: Stephane 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 [77] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles. When the covariate is functional and random (random design) we focus on kernel methods [18].

Conditional extremes are studied in climatology where one is interested in how climate change over years might affect extreme temperatures or rainfalls. In this case, the covariate is univariate (time). Bivariate examples include the study of extreme rainfalls as a function of the geographical location. The application part of the study is joint work with the LTHE (Laboratoire d'étude des Transferts en Hydrologie et Environnement) located in Grenoble [31], [32].

7.2.3. Estimation of extreme risk measures

Participant: Stephane Girard.

Joint work with: A. Daouia (Univ. Toulouse), L. Gardes (Univ. Strasbourg) and G. Stupfler (Univ. Aix-Marseille).

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 [64], 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 [63]. 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 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. Multivariate extremal events

Participants: Stephane Girard, Florence Forbes.

Joint work with: F. Durante (Univ. Bolzen-Bolzano, Italy) and G. Mazo (Univ. Catholique de Louvain, Belgique).

Copulas are a useful tool to model multivariate distributions [83]. However, while there exist various families of bivariate copulas, much fewer has been done when the dimension is higher. To this aim an interesting class of copulas based on products of transformed copulas has been proposed in the literature. The use of this class for practical high dimensional problems remains challenging. Constraints on the parameters and the product form render inference, and in particular the likelihood computation, difficult. As an alternative, we proposed a new class of copulas constructed by introducing a latent factor. Conditional independence with respect to this factor and the use of a nonparametric class of bivariate copulas lead to interesting properties like explicitness, flexibility and parsimony. In particular, various tail behaviours are exhibited, making possible the modeling of various extreme situations [17], [22].

7.2.5. Level sets estimation

Participant: Stephane Girard.

Joint work with: G. Stupfler (Univ. Aix-Marseille).

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 in collaboration with G. Stupfler that the so-called geometric quantiles have very counter-intuitive properties in such situations [20] and thus should not be used to detect outliers.

7.2.6. Robust Sliced Inverse Regression.

Participants: Stephane Girard, Alessandro Chiancone, Florence Forbes.

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

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 [14], we start considering 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.

7.2.7. Collaborative Sliced Inverse Regression.

Participants: Stephane Girard, Alessandro Chiancone.

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

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

In his PhD thesis work, Alessandro Chiancone studies the extension of the SIR method to different sub-populations. The idea is to assume that the dimension reduction subspace may not be the same for different clusters of the data [15]. One of the difficulty 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 clusterizes 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.

7.2.8. Hapke's model parameter estimation from photometric measurements

Participants: Florence Forbes, Emeline Perthame.

Joint work with: Sylvain Douté (IPAG, Grenoble).

The Hapke's model is a widely used analytical model in planetology to describe the spectro-photometry of granular materials. It is a non linear model F that links a set of parameters x to a "theoretical" Bidirectional Reflectance Diffusion Function (BRDF). In practice, we assume that the observed BRDF Y is a noisy version of the "theoretical" one

$$Y = F(x) + \epsilon \tag{5}$$

where ϵ is a centered Gaussian noise with diagonal covariance matrix Σ. Then x is also assumed to be random with some prior distribution to be specified, e.g. uniform on the parameters range in [84]. The overall goal is to estimate the posterior distribution $p(x|y)$ for some observed BRDF y. Equation (5) defines the likelihood of the model which is $p(y|x) = \mathcal{N}(y; F(x), \Sigma)$. Then since F is non linear, it is not possible to obtain an analytical expression for $p(x|y)$. However, it is easy to simulate parameters x that follows the posterior distribution $p(x|y) \propto p(y|x)\ p(x)$ for instance using MCMC techniques [84]. If only point estimate are desired, the MAP can be used and evolutionary algorithms can then be used also using $p(y|x)\ p(x)$ as a fitness function. But obtaining such simulations is time consuming and has to be done for each observed value of y. In this work, we propose to use a locally linear mapping approximation and an inverse regression strategy to provide an analytical expression of $p(x|y)$. The idea is that the non linear F can be approximated by a number K of locally linear functions and that each of this function is easy to inverse. It follows that the inverse of F is also approximated as locally linear. Preliminary results were presented at the MultiPlaNet workshop in Orsay, December 14, 2016. They show that the proposed method does not fully reproduce the previous results obtained using MCMC techniques. Further investigations are required to understand the origin of the difference. Also ABC (approximate Bayes computation) methods will be considered as a subsequent step that may improved the current procedure while remaining computationally efficient.

7.2.9. Prediction intervals for inverse regression models in high dimension

Participant: Emeline Perthame.

Joint work with: Emilie Devijver (KU Leuven, Belgium).

Inverse regression, as a dimension reduction technique, is a reliable and efficient approach to handle large regression issues in high dimension, when the number of features exceeds the number of observations. Indeed, under some conditions, dealing with the inverse regression problem associated to a forward regression problem drastically reduces the number of parameters to estimate and make the problem tractable. However, regression models are often used to predict a new response from a new observed profile of covariates, and we may be interested in deriving confidence bands for the prediction to quantify the uncertainty around a predicted response. Theoretical results have already been derived for the well-known linear model, but recently, the curse of dimensionality has increased the interest of practitioners and theoreticians into generalization of those results on a high-dimension context. When both the responses and the covariates are multivariate, we derive in this work theoretical prediction bands for the inverse regression linear model and propose an analytical expression of these intervals. The feasibility, the confidence level and the accuracy of the proposed intervals are also analyzed through a simulation study.

7.2.10. Multi sensor fusion for acoustic surveillance and monitoring

Participants: Florence Forbes, Jean-Michel Becu.

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

In the context of the DGA-rapid WIFUZ project with the ACOEM company, we addressed the issue of determining the localization of shots from multiple measurements coming from multiple sensors. We used Bayesian inversion and simulation techniques to recover multiple sources mimicking collaborative interaction between several vehicles. 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.

7.3. Graphical and Markov models

7.3.1. Conditional independence properties in compound multinomial distributions

Participant: Jean-Baptiste Durand.

Joint work with: Pierre Fernique (Inria, Virtual Plants) and Jean Peyhardi (Université de Montpellier).

We developed a unifying view of two families of multinomial distributions: the singular – for modeling univariate categorical data – and the non-singular – for modeling multivariate count data. In the latter model, we introduced sum-compound multinomial distributions that encompass re-parameterization of non-singular multinomial and negative multinomial distributions. The estimation properties within these compound distributions were obtained, thus generalizing know results in univariate distributions to the multivariate case. These distributions were used to address the inference of discrete-state models for tree-structured data. In particular, they were used to introduce parametric generation distributions in Markov-tree models [66].

7.3.2. Change-point models for tree-structured data

Participant: Jean-Baptiste Durand.

Joint work with: Pierre Fernique (Inria) and Yann Guédon (CIRAD), Inria Virtual Plants.

In the context of plant growth modelling, methods to identify subtrees of a tree or forest with similar attributes have been developed. They rely either on hidden Markov modelling or multiple change-point approaches. The latter are well-developed in the context of sequence analysis, but their extensions to tree-structured data are not straightforward. Their advantage on hidden Markov models is to relax the strong constraints regarding dependencies induced by parametric distributions and local parent-children dependencies. Heuristic approaches for change-point detection in trees were proposed and applied to the analysis of patchiness patterns (consisting of canopies made of clumps of either vegetative or flowering botanical units) in mango trees [43].

7.3.3. 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: Marianne Clausel (LJK) 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 make some kinds of press reviews. 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. 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.

To begin with, we first proved why HMM would be the best option in order to conduct this analysis and what could be the alternatives. A brief state of the art was made on models similar to HMMs. However, since our data is very specific, we needed to make use of unsupervised graphical generative models for the analysis of sequences which would keep a deep meaning. It resulted that Hidden semi-Markov model (HSMM) was the most powerful tool satisfying all our needs. Indeed, a HSMM is characterized by meaningful parameters such as an initial distribution, transition distributions, emission distributions and sojourn distributions, which allows us to directly characterize a reading strategy. Second, we found and improved an existing implementation of such a model. After searching for libraries to make inference in HSMM, the Vplants library embedded in the OpenAlea software turned out to be the most viable solution regarding the functionalities, though it was still incomplete. Consequently, we proposed improvements to this library and added functions in order to boost the likelihood of the data. This lead us to also propose a new library included in that software which is specific at the analysis of eye movements. Third, in order to improve and validate the interpretation of the reading strategies, we calculated indicators specific to each reading strategy. Fourth, since the parameters obtained from the model suggested individual and text variability, we first investigated text clustering to reduce the variability of the model. In order to do this, we supervised a group of 6 students to explore the text clustering component with the mission of clustering the texts by evolution of the semantic similarity throughout text. We therefore explored different methods for time series clustering and we retained the usage of Ascendant Hierarchical Clustering (AHC) using the Dynamic Time Warping (DTW) metric, which allows global dynamics of the time series to be captured, but not local dynamics. Plus, we preferred the simplicity and good understanding of the results using that method. Therefore, we deduced three text profiles giving meaning to the evolution of the semantic similarity: a step profile, a ramp profile, and a saw profile. With that new information in hand, we are now able to decompose our model over text profiles and hence, reduce its variability.

As discussed in the previous section, our work is focused on the standalone analysis of the eye-movements. We are currently polishing this phase of work. The common work and the 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.

7.3.4. Lossy compression of tree structures

Participant: Jean-Baptiste Durand.

Joint work with: Christophe Godin (Inria, Virtual Plants) and Romain Azais (Inria BIGS)

In a previous work [79], a method to compress tree structures and to quantify their degree of self-nestedness was developed. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph), equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the lossless compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree. A lossy version of the algorithm consists in coding the nearest self-nested tree embedded in the initial tree. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We obtained new theoretical results on the combinatorics of self-nested structures [60]. We improved this lossy compression method by computing a self-nested reduction of a tree that better approximates the initial tree. The algorithm has polynomial time complexity for trees with bounded outdegree. This approximation relies on an indel edit distance that allows (recursive) insertion and deletion of leaf vertices only. We showed using a simulated dataset that the error rate of this lossy compression method is always better than the loss based on the nearest embedded self-nestedness tree [79] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. Perspectives of improving the time complexity of our algorithm include taking profit from one of its byproduct, which could be used as an indicator of both the number of potential candidates to explore and of the proximity of the tree to the nearest self-nested tree.

7.3.5. Learning the inherent probabilistic graphical structure of metadata

Participants: Thibaud Rahier, Stephane Girard, Florence Forbes.

Joint work with: Sylvain Marié, Schneider Electric.

The quality of prediction and inference on temporal data can be significantly improved by taking advantage of the associated metadata. However, metadata are often only partially structured and may contain missing values. In the context of T. Rahier's PhD with Schneider Electric, we first considered the problem of learning the inherent probabilistic graphical structure of metadata, which has two main benefits: (i) graphical models are very flexible and therefore enable the fusion of different types of data together (ii) the learned graphical model can be interrogated to perform tasks on metadata alone: variable clustering, conditional independence discovery or missing data replenishment. Bayesian Network (and more generally Probabilistic Graphical Model) structure learning is a tremendous mathematical challenge, that involves a NP-Hard optimisation problem. In the past year, we have explored many approaches to tackle this issue, and begun to develop a tailor-made algorithm, that exploits dependencies typically present in metadata, and that significantly speeds up the structure learning task and increases the chance of finding the optimal structure.

7.3.6. Robust Graph estimation

Participants: Karina Ashurbekova, Florence Forbes.

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

In the face of increasingly high dimensional data and of trying to understand the dependency/association present in the data the literature on graphical modelling is growing rapidly and covers a range of applications (from bioinformatics e.g gene expression data to document modelling). A major limitation of recent work on using the (standard) Student t distribution for robust graphical modelling is the lack of independence and conditional independence of the Student t distribution, and estimation in this context (with the standard student t) is very difficult. We propose to develop and assess a generalized Student t from a new family (which has independence and conditional independence as special properties) for the general purpose of graphical modelling in high dimensional settings. Its main characteristic is to include multivariate heavy-tailed distributions with variable marginal amounts of tailweight that allow more complex dependencies than the standard case. We target an application to brain connectivity data for which standard Gaussian graphical models have been applied. Brain connectivity analysis consists in the study of multivariate time series representing local dynamics at each of multiple sites or sources throughout the whole human brain while functioning using for example functional magnetic resonance imaging (fMRI). The acquisition is difficult and often spikes are observed due to the movement of the subjects inside the scanner. In the case of identifying Gaussian graphical models, the glasso technique has been developed for estimating sparse graphs. However, this method can be severely impacted by the inclusion of only a few contaminated values, such as spikes that commonly occur in fMRI time series, and the resulting graph has the potential to contain false positive edges. Therefore, our goal was to assess the performance of more robust methods on such data.

7.4. Robust non Gaussian models

7.4.1. Robust Locally linear mapping with mixtures of Student distributions

Participants: Florence Forbes, Emeline Perthame, Brice Olivier.

The standard GLLiM model [6] for high dimensional regression assumes Gaussian noise models and is in its unconstrained version equivalent to a joint GMM. The fact that response and independent variables (X, Y) are jointly a mixture of Gaussian distribution is the key for all derivations in the model. In this work, we show that similar developments are possible based on a joint Student Mixture model, joint SMM. It follows a new model referred to as SLLiM for Student Locally linear mapping for which we investigate the robustness to outlying data in a high dimensional regression context [71]. The corresponding code is available on the CRAN in the *xLLiM* package.

7.4.2. Rectified binaural ratio: A complex T-distributed feature for robust sound localization

Participant: Florence Forbes.

Joint work with: Antoine Deleforge, Inria PANAMA team in Rennes.

Most existing methods in binaural sound source localization rely on some kind of aggregation of phase-and level-difference cues in the time-frequency plane. While different aggregation schemes exist, they are often heuristic and suffer in adverse noise conditions. In this work, we introduce the rectified binaural ratio as a new feature for sound source localization. We show that for Gaussian-process point source signals corrupted by stationary Gaussian noise, this ratio follows a complex t-distribution with explicit parameters. This new formulation provides a principled and statistically sound way to aggregate binaural features in the presence of noise. We subsequently derive two simple and efficient methods for robust relative transfer function and time-delay estimation. Experiments on heavily corrupted simulated and speech signals demonstrate the robustness of the proposed scheme. This work has been presented at the Eusipco conference in 2016 [30].

7.4.3. Statistical reconstruction methods for multi-energy tomography

Participants: Florence Forbes, Pierre-Antoine Rodesch.

Joint work with: Veronique Rebuffel from CEA 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.

7.5. Statistical models for Neuroscience

7.5.1. Advanced statistical analysis of functional Arterial Spin Labelling data

Participants: Florence Forbes, Aina Frau Pascual.

Joint work with: Philippe Ciuciu from Team PARIETAL and Neurospin, CEA Saclay.

Arterial Spin Labelling (ASL) is a non-invasive perfusion MR imaging technique that can be also used to measure brain function (fASL for functional ASL). In contrast to BOLD fMRI, it gives a quantitative and absolute measure of cerebral blood flow (CBF), making this modality appealing for clinical neuroscience and patient's follow-up over longitudinal studies. However, its limited signal-to-noise ratio makes the analysis of fASL data challenging. In this work, we compared different approaches (GLM vs JDE) in the analysis of functional ASL data for the detection of evoked brain activity at the group level during visual and motor task performance. Our dataset has been collected at Neurospin on a 3T Tim Trio Siemens scanner (CEA Saclay, France), during the HEROES project (Inria Grant). It contains BOLD data (165 scans, TR=2.5s, TE=30ms, 3x3x3mm3) and functional pulsed ASL data (Q2TIPS PICORE scheme [Luh,00], 165 scans, TR=2.5s, TE=11ms, 3x3x7.5 mm3) of 13 right-handed subjects (7 men and 6 women) of age between 20 and 29. The experimental design consists of a mini-block paradigm of visual, motor and auditory tasks with 16 blocks of 15s each followed by 10s of rest. Data have been scaled, realigned, and normalized. For univariate analysis, the images have also been spatially smoothed with a Gaussian kernel of 5 mm full width half at maximum. Three data analysis approaches have been compared: (a) univariate General Linear Model (GLM) that considers canonical shapes for the perfusion and hemodynamic responses; (b) physiologically informed joint detection estimation (PI-JDE) [4] that jointly estimates effect maps and response functions in a multivariate manner in a Bayesian framework; (c) A restricted version of PI-JDE that considers fixed canonical shapes for the perfusion and hemodynamic responses (PRF and HRF, respectively), defining an intermediate approach between the first two. Since methods (b)-(c) embed adaptive spatial regularization, they do not require a preliminary smoothing of the data. Our results demonstrate that the PI-JDE multivariate approach is a competing alternative to GLM for the analysis of fASL: it recovers more localized and stronger effects. Our findings also replicate the state-of-the-art by showing more localized activation patterns in perfusion as compared to hemodynamics.

7.5.2. BOLD VEM multi session extension of the JDE approach

Participants: Florence Forbes, Aina Frau Pascual.

Joint work with: Philippe Ciuciu from Team PARIETAL and Neurospin, CEA Saclay.

The fast solution of the JDE approach for BOLD fMRI presented in [5] uses a variational expectation maximization (VEM) algorithm and considers a single session of BOLD data. This paper shows the faster performance of this algorithm with respect to the Markov Chain Monte Carlo (MCMC) approach presented in earlier work, with similar results. In fMRI, usually several sessions are acquired for the same subject to be able to compare them or combine them. In [73], a multiple-session extension of the JDE approach has been proposed to analyze several sessions together. The solution proposed uses MCMC and considers that the response levels have a mean value per condition and a common variance between sessions. In the context of Aina Frau's PhD, a VEM solution of this extension has been implemented. Experimental results have shown that the solution of the multiple-session VEM is not very different from the average of the results computed with single session VEM. For this reason, we proposed a heteroscedastic version of the multiple-session VEM. It amounts to considering session-specific variances. The goal is to be able to weight the importance of the different sessions so as to diminish the contribution of any potential noisy session to the final parameter estimates.

7.5.3. Estimating biophysical parameters from multimodal fMRI data

Participants: Florence Forbes, Pablo Mesejo Santiago.

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

Functional Magnetic Resonance Imaging (fMRI) indirectly studies brain function. With Jan M. Warnking (Grenoble Institute of Neurosciences) we worked on the estimation of biophysical parameters from fMRI signals. We first used only BOLD signals, using a stochastic population-based optimization method to estimate 15 parameters without neither providing initial estimates nor computing gradients. Initial results were published at MICCAI 2015 and in the IEEE JSTSP journal [81], [82]. Also a MATLAB toolbox was released (see software section). The current ongoing work is to study the impact of the combination of different fMRI modalities in the estimation of this biophysical parameters. We can use 3 fMRI modalities (BOLD, ASL and MION) and 13 rats. We ran our optimizer with all possible combinations of modalities. The initial hypothesis was that as long as we introduce more fMRI modalities we would like to see more consistent estimates but we need to assess possible limits due to potential lack of data: only 13 rats, 6 of them without MION, and potential outliers among the rats that would better be excluded from the analysis.

7.5.4. Multi-subject joint parcelation detection estimation in functional MRI

Participant: Florence Forbes.

Joint work with: Lotfi Chaari, Mohanad Albughdadi, Jean-Yves Tourneret from IRIT-ENSEEIHT in Toulouse and Philippe Ciuciu from Neurospin, CEA Saclay.

fMRI experiments are usually conducted over a population of interest for investigating brain activity across different regions, stimuli and subjects. Multi-subject analysis usually proceeds in two steps: an intra-subject analysis is performed sequentially on each individual and then a group-level analysis is carried out to report significant results at the population level. This work considers an existing Joint Parcellation Detection Estimation (JPDE) model which performs joint hemodynamic parcellation, brain dynamics estimation and evoked activity detection. The hierarchy of the JPDE model is extended for multi-subject analysis in order to perform group-level parcellation. Then, the corresponding underlying dynamics is estimated in each parcel while the detection and estimation steps are iterated over each individual. Validation on synthetic and real fMRI data shows its robustness in inferring group-level parcellation and the corresponding hemodynamic profiles. This work has been presented at ISBI 2016 [42].

7.5.5. Automatic segmentation and characterization of brain tumors using robust multivariate clustering of multiparametric MRI

Participants: Florence Forbes, Alexis Arnaud.

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

Brain tumor segmentation is a difficult task in the field of multiparametric MRI analysis because of the number of maps that are available. Furthermore, the characterization of brain tumors can be time-consuming, even for medical experts, and the reference method is biopsy which is a local and invasive technique. Because of this, it is important to develop automatic and non-invasive approaches in order to help the medical expert with these issues. In this study we use a robust statistical model-based method to classify multiparametric MRI of rat brains. The voxels are gather into classes resulting from multivariate multi-scaled Student distributions, which can accommodate outliers. First we adjust a mixture model on a reference group of rats to learn the MRI characteristics of healthy tissues. Second we use this model to delineate the brain tumors as atypical voxels in the data set of unhealthy rats. Third we adjust a new mixture model only on the atypical voxels to learn the MRI characteristics of tumorous tissues. Finally, we extract a fingerprint for each tumor type to make a tumor dictionary.

Our data set is composed of healthy rats (n=8 rats) and 4 groups of rats bearing a brain tumor model (n=8 per group). For each rat, we acquired 5 quantitative MRI parameters along 5 slices. And the proposed tumor dictionary reaches a rate of 75% of accurate prediction with a leave-one-out procedure.

7.5.6. Monitoring brain tumor evolution using multiparametric MRI

Participants: Florence Forbes, Alexis Arnaud.

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

Analyzing brain tumor tissue composition can 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 were 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.

7.5.7. Assessment of tissue injury in severe brain trauma

Participant: Florence Forbes.

Joint work with: Michel Dojat and Christophe Maggia from Grenoble Institute of Neuroscience and Senan Doyle from Pixyl.

Traumatic brain injury (TBI) remains a leading cause of death and disability among young people worldwide and current methods to predict long-term outcome are not strong. TBI initiates a cascade of events that can lead to secondary brain damage or exacerbate the primary injury, and these develop hours to days after the initial accident. The concept of secondary brain damage is the focus of modern TBI management in Intensive Care Units. The imbalance between oxygen supply to the brain tissue and utilization, i.e. brain tissue hypoxia, is considered the major cause for the development of secondary brain damage, and hence poor neurological outcome. Monitoring brain tissue oxygenation after TBI using brain tissue O_2 pressure (PbtO_2) probes surgically inserted into the parenchyma, may help clinicians to initiate adequate actions when episodes of brain ischemia/hypoxia are identified. The aggressive treatment of low PbtO_2 values (< 15mmHg for more than 30 minutes) was associated with better outcome compared to standard therapy in some cohort studies of severe head-injury patients. However, another study was unable to find similar benefits to patient outcome. MRI is an excellent modality for estimating global and regional alterations in TBI and for following their longitudinal evolution. To assess the complexity of TBI, several morphological sequences are required for assessing volume

loss. Moreover, diffusion tensor imaging (DTI) offers the most sensitive modality for the detection of changes in the acute phase of TBI and increases the accuracy of long-term outcome prediction compared to the available clinical/radiographic prognostic score. Mean Diffusivity (MD) or Apparent Diffusion Coefficient (ADC) have been widely used to determine the volume of ischemic tissue, and assess intra- and extracellular conditions. A reduction of MD is related to cytotoxic edema (intracellular) while an increase of MD indicates a vasogenic edema (extracellular). Changes of MD are expected with severe TBI. The volume of lesions on DTI shows a strong correlation with neurological outcome at patient discharge. We consider a clinically relevant criterion to be the volume of vulnerable brain lesions after TBI, as previously suggested. In consequence, we need an automatic segmentation method to assess the tissue damage in severe trauma, acute phase i.e. before 10 days after the event. Skull deformation, the presence of blood in the acute phase, the high variability of brain damage that excludes the use of anatomical *a priori* information, and the diffuse aspect of brain injury affecting potentially all brain structures, render TBI segmentation particularly demanding. The methods proposed in the literature are mainly concerned with volumetric changes following TBI and scarcely report lesion load. In this work, we report our methodological developments to assess lesion load in severe brain trauma in the entire brain. We use P-LOCUS to perform brain tissue segmentation and exclude voxels labeled as CSF, ventricle and hemorrhagic lesion. We propose a fusion of several atlases to parcel cortical, subcortical and WM structures into well identified regions where MD values can be expected to be homogenous. Abnormal voxels are detected in these regions by comparing MD values with normative values computed from healthy volunteers. The preliminary results, evaluated in a single center, are a first step in defining a robust methodology intended to be used in multi-center studies. This work has been published in [58].

7.5.8. Automatic multiple sclerosis lesion segmentation with P-Locus

Participant: Florence Forbes.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience and Senan Doyle from Pixyl.

P-LOCUS provides automatic quantitative neuroimaging biomarker extraction tools to aid diagnosis, prognosis and follow-up in multiple sclerosis studies. The software performs accurate and precise segmentation of multiple sclerosis lesions in a multi-stage process. In the first step, a weighted Gaussian tissue model is used to perform a robust segmentation. The algorithm avails of complementary information from multiple MR sequences, and includes additional estimated weight variables to account for the relative importance of each voxel. These estimated weights are used to define candidate lesion voxels that are not well described by a normal tissue model. In the second step, the candidate lesion regions are used to populate the weighted Gaussian model and guide convergence to an optimal solution. The segmentation is unsupervised, removing the need for a training dataset, and providing independence from specific scanner type and MRI scanner protocol. The procedure was applied to participate to the MSSEG Challenge at Miccai 2016 in Athen: Multiple Sclerosis Lesions Segmentation Challenge Using a Data Management and Processing Infrastructure [55].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

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

PhD contract with EDF (2016-2018). S. Girard is the advisor of a PhD (A. Clément) with EDF. The goal is to investigate sensitivity analysis and extrapolation limits in Extreme value theory with application to river flows analysis.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- 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 P. Mesejo are co-organizing a reading group on Deep Learning with R. Horaud and K. Alahari.

9.2. National Initiatives

9.2.1. Grenoble Idex projects

MISTIS is involved in a newly accepted transdiciplinary project **NeuroCoG** (December 2016). F. Forbes is also responsible for a workpackage in another project entitled "Institut des sciences des données".

9.2.2. Competitivity Clusters

The MINALOGIC VISION 4.0 project:MISTIS is involved in a new (October 2016) three-year *Pôle de competitivité Minalogic* project. The project is led by VI-Technology (http://www.vitechnology.com), a world leader in Automated Optical Inspection (AOI) of a broad range of electronic components. The other partners are the G-Scope Lab in Grenoble and ACTIA company based in Toulouse. Our goal is to exploit more intensively statistical techniques to exploit the large amount of data registered by AOI machines.

9.2.3. Defi Mastodons CNRS

Defi La qualité des données dans le Big Data (2016-17). S. Girard is involved in a 1-year project entitled “Classification de Données Hétérogènes avec valeurs manquantes appliquée au Traitement des Données Satellitaires en écologie et Cartographie du Paysage”, the other partners being members of Modal (Inria Lille Nord-Europe) or ENSAT-Toulouse. The total funding is 10 keuros.

9.2.4. Defi Imag’IN CNRS

Defi Imag’IN MultiPlanNet (2015-2016). 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.

9.2.5. 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).

9.2.6. 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.

9.3. International Initiatives

9.3.1. Inria International Labs

LIRIMA

Associate Team involved in the International Lab:

9.3.1.1. SIMERGE

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

International Partner (Institution - Laboratory - Researcher):

UGB (Senegal) - LERSTAD - Abdou Kâ Diongue

Starting year: 2015

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

The objective of the associate team is to federate some researchers from LERSTAD (Laboratoire d'Etudes et de Recherches en Statistiques et Développement, Université Gaston Berger) and MISTIS (Inria Grenoble Rhône-Alpes). The associate team will consolidate the existing collaborations between these two laboratories. Since 2010, the collaborations have been achieved through the co-advising of two PhD theses. They have led to three publications in international journals. The associate team will also involve 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. We aim at developing two research themes: 1) Spatial extremes with application to management of extreme risks and 2) Classification with application to global epidemiology.

9.3.1.2. 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 Washington in Seattle, Université Gaston Berger in Senegal and Universities of Melbourne and Brisbane in Australia. In 2016, new collaborations had started with the statistics department of University of Michigan, in Ann Arbor, USA and with the statistics department of McGill University in Montreal, Canada.

The main active international collaborations in 2016 are with:

- F. Durante, Free University of Bozen-Bolzano, Italy.
- K. Qin and D. Wraith resp. from RMIT 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.
- A. Nazin from Russian Academy of Science in Moscow, Russia.
- 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.

9.3.2. Participation in Other International Programs

Alexis Arnaud received an award from the MITACS program, for a 5 months visit to McGill university in Montreal.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Seydou Nourou Sylla (Université Gaston Berger, Sénégal) has been hosted by the MISTIS team for two months.
- Naisyin Wang and Chun-Chen Tu from University of Michigan, Ann Arbor, USA, have been hosted by the MISTIS team for one week.

9.4.2. Visits to International Teams

S. Girard went to univ. Gaston Berger in St Louis Senegal in the context of the SIMERGE associated team.

9.4.2.1. Research Stays Abroad

Alexis Arnaud spent 5 months at McGill university in Montreal.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Stéphane Girard co-organized the workshop "Learning with functional data", held in Lille (October 2016), He was also a member of the organizing committee of “Journées MAS de la SMAI”, held in Grenoble (August 2016)
- Florence Forbes and Stéphane Girard co-organized a session “Dimension reduction for regression” at the ERCIM conference in Séville, Spain (December 2016).
- Julyan Arbel organized the Bayesian nonparametric prediction session at the International Society of Bayesian Analysis Conference, June 2016. He also co-organized the StaTalk Workshop on Bayesian nonparametrics, Collegio Carlo Alberto, Moncalieri, Italy, February 19.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Stéphane Girard was the president of the scientific committee of the CIMPA conference “Méthodes statistiques pour l’évaluation des risques extrêmes”, held in Saint-Louis, Sénégal (April 2016), .
- Stéphane Girard was a member of the conference program committee of the “Mathematical Finance and Actuarial Sciences conference” organized by the AIMS (African Institute for Mathematical Sciences), Mbour, Sénégal (July 2016).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Stéphane Girard is Associate Editor of the *Statistics and Computing* journal since 2012 and Associate Editor of the *Journal of Multivariate Analysis* since 2016. He was co-editor of the book *Statistics for astrophysics, clustering and classification*, vol. 77, EDP sciences, 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.

10.1.3.2. Reviewer - Reviewing Activities

In 2016, S. Girard has been a reviewer for *Australian and New Zealand Journal of Statistics, Extremes* and *Dependence Modelling.*

In 2016, F. Forbes has been a reviewer for *Journal of Multivariate Analysis, Statistics and Computing, Computational Statistics and Data Analysis* .

In 2016, Julyan Arbel has been reviewer for NIPS 2016, ICML 2016, AISTATS 2016, the Annals of Statistics, Bayesian Analysis, Bernoulli, Biometrics, the Canadian Journal of Statistics, the Hacettepe Journal of Mathematics and Statistics, the Journal of Agricultural, Biological, and Environmental Statistics, SoftwareX, Statistics and Computing.

10.1.4. Invited Talks

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

- "Extreme value modeling and water ressources" workshop (Aussois) [29],

- 3rd conference of the International Society for Non-Parametric Statistics (Avignon) [37],

- "Extremes - Copulas - Actuarial sciences" workshop (Luminy) [40],

- Statistics workshop at Tilburg University (Netherlands)

- ERCIM CFE-CMStatistics (Seville, Spain) [39].

Florence Forbes has been invited to give talks at :

- the 23th summer session of the Working Group on Model-based Clustering, Paris, July 18-23, 2016.

- the Fédération Rhône-Alpes-Auvergne day on multivariate data analysis in Grenoble, October 2016,

- the 11th Peyresq summer school on signal and image processing (July 2016): 5 hour lecture on Bayesian Analysis and applications [67].

- a special session on Dimension reduction for regression at the ERCIM CFE-CMStatistics conference, December 2016, in Seville, Spain [35],

- at the annual meeting of the MultiPlaNet project (Defi Imag'In CNRS) in Orsay in December 2016, on the inversion of the Hapke's model from photometric measurements.

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

- Rencontres Statistiques Lyonnaises, Institut Camille Jordan, November 23. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Séminaire de Statistique du LJK, Université Grenoble Alpes, November 17. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- The Bayes Club, Korteweg-de Vries Institute for Mathematics, University of Amsterdam, October 7. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Séminaire de Statistique, Université Lille 3, October 6. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Séminaire de Proba-Stat, Université Paris 12 Créteil, October 4. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Séminaire de Proba-Stat, Université de Franche-Comté, Besancon, September 5. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- ISBA World Meeting, Sardinia, Italy, June 13-17. Invited talk: Bayesian nonparametric inference for discovery probabilities.

- Mistis Seminar, Inria Grenoble, France, February 12. Talk: Infinite mixture models in Bayesian nonparametrics.

Julyan Arbel presented at the following contributed sessions in conferences and workshops:

- NIPS Meeting, Barcelona, Spain, Poster: Truncation error of a superposed gamma process in a decreasing order representation, Poster: Advances in Approximate Bayesian Inference workshop, Dec 9.

- NIPS Meeting, Barcelona, Spain, Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models, Practical Bayesian Nonparametrics workshop, Dec 9.

- Journées MAS, Grenoble, France, August 29-31. Poster: Bayesian nonparametrics, why and how?

- Third Bayesian Young Statisticians Meeting, Florence, Italy, June 19-21. Talk: A moment-matching Ferguson & Klass algorithm.

- Journées de Statistique de la SFdS, Montpellier, France, May 30 - June 3. Talk: Bayesian nonparametric inference for discovery probabilities.

- StaTalk Workshop, Collegio Carlo Alberto, Moncalieri, Italy, February 19. Talk 1: A gentle introduction to Bayesian Nonparametrics. Talk 2: Species sampling models.

- MCMSki V, Lenzerheide, Switzerland, January 5-7. Invited talk: A moment-matching Ferguson & Klass algorithm.

Emeline Perthame has been invited to give a talk at:

- the statistics seminar at University of Caen in October 2016 on an *Inverse regression approach to robust non-linear high-to-low dimensional mapping*.

Gildas Mazo has been invited to give a talk at:

- a special session on Copulas at the ERCIM CFE-CMStatistics conference, December 2016, in Seville, Spain.

Alexis Arnaud gave a talk at:

- a GdR ISIS meeting on *Méthodes d'apprentissage statistiques et applications à la santé* 2016-10-21, Telecom Paris, on Automatic segmentation and characterization of brain tumors using robust multivariate clustering of multiparametric MRI.

10.1.5. Leadership within the Scientific Community

Stéphane Girard is at the head of the associate team (*Statistical Inference for the Management of Extreme Risks and Global Epidemiology*) created in 2015 between IISTIS 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 in charge of evaluating research projects for the Research Foundation Flanders (FWO), Belgium.
- Stéphane Girard is a Voting Member for the International Society for NonParametric Statistics (ISNPS).

10.1.7. Research Administration

- Stéphane Girard has been at the head of the Probability and Statistics department of the LJK (Laboratoire Jean Kuntzmann) from September 2012 to September 2016.
- 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.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : Stéphane Girard, *Statistique Inférentielle Avancée*, 41 ETD, M1 level, Ensimag. Grenoble-INP, France.

Master : Stéphane Girard, *Introduction à la statistique des valeurs extrêmes*, 15 ETD, M2 level, Université Gaston Berger, Saint-Louis, Sénégal.

Licence : Alexis Arnaud, *Probability and statistics*, 56 ETD, L2 level, IUT2 Grenoble, Université Pierre Mendès France.

Master: Jean-Baptiste Durand, *Statistics and probability*, 192 ETD, M1 and M2 levels, Ensimag Grenoble INP, France. Head of the MSIAM M2 programme, in charge of the statistics and data science tracks ([12]).

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

J-M. Becu, C. Albert, B. Olivier are teaching at UGA.

Master and PhD course: Julyan Arbel gave a course on Bayesian statistics, 30 ETD, Collegio Carlo Alberto, Moncalieri, Turin, Italy.

10.2.2. Supervision

Aina Frau-Pascual, "*Statistical models for the analysis of ASL and BOLD functional magnetic resonance modalities to study brain function and disease*", defended on December 19, 2016, Université Grenoble-Alpes, supervised by Florence Forbes and Philippe Ciuciu (CEA, Inria PARIETAL).

Seydou Nourou Sylla, "*Modélisation et classification de données binaires en grande dimension - Application à l'autopsie verbale*", defended on December 21, 2016, Université Gaston Berger, Saint-Louis, Sénégal, supervised by Abdou Diongue (Université Gaston Berger, Sénégal) and Stéphane Girard.

Alessandro Chiancone, "*Réduction de dimension via Sliced Inverse Regression: Idées et nouvelles propositions*", defended on October 28, 2016, Université Grenoble-Alpes, supervised by Stéphane Girard and Jocelyn Chanussot (Grenoble INP).

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: Maïlys Lopes, "*Télédétection en écologie du paysage : statistiques en grande dimension pour la multirésolution spatiale et la haute résolution temporelle*", started on November 2014, Stéphane Girard and Mathieu Fauvel (INRA Toulouse).

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

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

10.2.3. Juries

10.2.3.1. PhD

- Stéphane Girard has been reviewer of three PhD theses in 2016:

  - Cees de Valk, "*A large deviation approach to the statistics of extreme events*", Tilburg University, Netherlands, December 2016.

  - Nicolas Goix, "*Apprentissage automatique et extrêmes et pour la détection d'anomalies*", Telecom ParisTech, november 2016.

  - Anthony Zullo, "*Analyse de données fonctionnelles en télédétection hyperspectrale : application à l'étude des paysages agri-forestiers*", Univ. Toulouse, September 2016.

- S. Girard was a member of two PhD committees in 2016:

  - Quentin Sebille, "*Modélisation spatiale de valeurs extrêmes, application à l'étude de précipitations en France*, Univ. Lyon, december 2016.

  - Khalil Said, "*Mesures de risque multivariées et applications en science actuarielle*", Univ. Lyon, december 2016.

- Florence Forbes has been reviewer of 1 PhD thesis in 2016:

  - Hong Phuong Dang, December 1st, 2016, Centrale Lille.

- F. Forbes was a member of one PhD committee in 2016:

  - Mohanad Albughdadi, September 2016, ENSHEEIT, Toulouse.

10.2.3.2. HDR

S. Girard was a member of the HDR committee of Mathieu Ribatet, Univ. Montpellier, November 2016.

F. Forbes was in the HDR committee of Sophie Achard, Univ. Grenoble Alpes, May 2016.

10.2.3.3. Other committees

- S. Girard is a member of the "Comité des Emplois Scientifiques" at Inria Grenoble Rhône-Alpes since 2015.
- 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.
- 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.
- F. Forbes has been a member of 2 selection committees for Professors at Centrale-Supelec and Centrale Nantes and 1 selection committee for Assistant Professor at Paris-Sud University.

10.3. Popularization

- S. Girard presented his research on extreme-value analysis at the “Conférence ISN et enseignement”, March 2016, video. He also gave a talk at the Institut de Maitrise des Risques (IMdR) [38] on a similar topic.
- Julyan Arbel led the Math en Jeans teams at Lycée francais Jean Giono, Turin, working on various subjects spanning from statistics, machine learning, to combinatorics and games.

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

Articles in International Peer-Reviewed Journals

[11] 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", 2017, https://hal.archives-ouvertes.fr/hal-01426385

[12] 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

[13] J. ARBEL, V. COSTEMALLE. *Estimation of immigration flows : reconciling two sources by a Bayesian approach*, in "Economie et Statistique", April 2016, https://hal.archives-ouvertes.fr/hal-01396606

[14] A. CHIANCONE, F. FORBES, S. GIRARD. *Student Sliced Inverse Regression*, in "Computational Statistics and Data Analysis", August 2016 [*DOI :* 10.1016/J.CSDA.2016.08.004], https://hal.archives-ouvertes.fr/hal-01294982

[15] A. CHIANCONE, S. GIRARD, J. CHANUSSOT. *Collaborative Sliced Inverse Regression*, in "Communication in Statistics - Theory and Methods", 2016, forthcoming [*DOI :* 10.1080/03610926.2015.1116578], https://hal.inria.fr/hal-01158061

[16] J.-B. DURAND, Y. GUÉDON. *Localizing the latent structure canonical uncertainty: entropy profiles for hidden Markov models*, in "Statistics and Computing", 2016, vol. 26, n^{o} 1, pp. 549-567, The final publication is available at Springer via http://dx.doi.org/10.1007/s11222-014-9494-9 [*DOI :* 10.1007/s11222-014-9494-9], https://hal.inria.fr/hal-01090836

[17] F. DURANTE, S. GIRARD, G. MAZO. *Marshall–Olkin type copulas generated by a global shock*, in "Journal of Computational and Applied Mathematics", April 2016, vol. 296, pp. 638–648 [*DOI :* 10.1016/J.CAM.2015.10.022], https://hal.archives-ouvertes.fr/hal-01138228

[18] L. GARDES, S. GIRARD. *On the estimation of the functional Weibull tail-coefficient*, in "Journal of Multivariate Analysis", 2016, vol. 146, pp. 29–45, https://hal.archives-ouvertes.fr/hal-01063569

[19] I. D. GEBRU, X. ALAMEDA-PINEDA, F. FORBES, R. HORAUD. *EM Algorithms for Weighted-Data Clustering with Application to Audio-Visual Scene Analysis*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", December 2016, vol. 38, n° 12, pp. 2402 - 2415 [*DOI :* 10.1109/TPAMI.2016.2522425], https://hal.inria.fr/hal-01261374

[20] S. GIRARD, G. STUPFLER. *Intriguing properties of extreme geometric quantiles*, in "REVSTAT - Statistical Journal", 2016, forthcoming, https://hal.inria.fr/hal-00865767

[21] P. JORDANOVA, Z. FABIÁN, P. HERMANN, L. STRELEC, A. RIVERA, S. GIRARD, S. TORRES, M. STEHLÍK. *Weak properties and robustness of t-Hill estimators*, in "Extremes", 2016, vol. 19, n° 4, pp. 591–626, https://hal.archives-ouvertes.fr/hal-01327002

[22] G. MAZO, S. GIRARD, F. FORBES. *A flexible and tractable class of one-factor copulas*, in "Statistics and Computing", September 2016, vol. 26, n° 5, pp. 965-979, https://hal.archives-ouvertes.fr/hal-00979147

[23] P. MESEJO, O. IBÁÑEZ, O. CORDÓN, S. CAGNONI. *A Survey on Image Segmentation using Metaheuristic-based Deformable Models: State of the Art and Critical Analysis*, in "Applied Soft Computing", April 2016, https://hal.archives-ouvertes.fr/hal-01282678

[24] P. MESEJO, D. PIZARRO, A. ABERGEL, O. ROUQUETTE, S. BEORCHIA, L. POINCLOUX, A. BARTOLI. *Computer-Aided Classification of Gastrointestinal Lesions in Regular Colonoscopy*, in "IEEE Transactions on Medical Imaging", 2016 [*DOI :* 10.1109/TMI.2016.2547947], https://hal.archives-ouvertes.fr/hal-01291797

[25] P. MESEJO, S. SAILLET, O. DAVID, C. BÉNAR, J. M. WARNKING, F. FORBES. *A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data*, in "IEEE Journal of Selected Topics in Signal Processing", March 2016, vol. 10, n° 2, pp. 416-427 [*DOI :* 10.1109/JSTSP.2015.2502553], https://hal.inria.fr/hal-01221115

[26] M. STEHLIK, P. AGUIRRE, S. GIRARD, P. JORDANOVA, J. KISEĽÁK, S. TORRES-LEIVA, Z. SADOVSKY, A. RIVERA. *On ecosystems dynamics*, in "Ecological Complexity", 2016, forthcoming, https://hal.inria.fr/hal-01394734

[27] W. YANG, B. PALLAS, J.-B. DURAND, S. S. MARTINEZ, M. HAN, E. COSTES. *The impact of long-term water stress on tree architecture and production is related to changes in transitions between vegetative and reproductive growth in the 'Granny Smith' apple cultivar*, in "Tree Physiology", September 2016 [*DOI :* 10.1093/TREEPHYS/TPW068], https://hal.inria.fr/hal-01377095

Invited Conferences

[28] A. DAOUIA, S. GIRARD, G. STUPFLER. *Estimation of the marginal expected shortfall using extreme expectiles*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics", Seville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01415581

[29] A. DAOUIA, S. GIRARD, G. STUPFLER. *Tail risk estimation based on extreme Lp-quantiles*, in "Workshop "Extreme value modeling and water ressources"", Aussois, France, 2016, https://hal.archives-ouvertes.fr/hal-01340767

[30] A. DELEFORGE, F. FORBES. *Rectified binaural ratio: A complex T-distributed feature for robust sound localization*, in "European Signal Processing Conference", Budapest, Hungary, August 2016, pp. 1257-1261, https://hal.inria.fr/hal-01372337

[31] J. EL METHNI, L. GARDES, S. GIRARD. *Estimation of risk measures for extreme pluviometrical measurements*, in "Workshop "Extreme value modeling and water ressources"", Aussois, France, 2016, https://hal.archives-ouvertes.fr/hal-01340774

[32] J. EL METHNI, L. GARDES, S. GIRARD. *Estimation of risk measures for extreme pluviometrical measurements*, in "26th Annual Conference of The International Environmetrics Society", Edimbourg, United Kingdom, July 2016, https://hal.archives-ouvertes.fr/hal-01350104

[33] J. EL METHNI, L. GARDES, S. GIRARD. *Frontier estimation based on extreme risk measures*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics", Seville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01415591

[34] J. EL METHNI, S. GIRARD, L. GARDES. *Kernel estimation of extreme risk measures for all domains of attraction*, in "Extremes, Copulas and Actuarial Sciences", Marseille, France, February 2016, https://hal.inria.fr/hal-01312846

[35] F. FORBES, A. CHIANCONE, S. GIRARD. *Student sliced inverse regression*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics", Seville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01415576

[36] F. FORBES, A. CHIANCONE, S. GIRARD. *Student Sliced Inverse Regression*, in "23th summer session of the Working Group on Model-based Clustering", Paris, France, July 2016, https://hal.archives-ouvertes.fr/hal-01423626

[37] L. GARDES, S. GIRARD. *Estimation of the functional Weibull-tail coefficient*, in "3rd conference of the International Society for Non-Parametric Statistics (ISNPS)", Avignon, France, June 2016, https://hal.inria.fr/hal-01366174

[38] S. GIRARD, C. ALBERT, A. DUTFOY. *Extrapolation dans les queues de distribution avec la théorie des valeurs extrêmes*, in "Journée estimation de probabilités d'événements rares en maitrise des risques et en sûreté de fonctionnement", Cachan, France, Institut de Maitrise des Risques (IMdR), 2016, https://hal.archives-ouvertes.fr/hal-01330131

[39] S. GIRARD, A. DAOUIA, G. STUPFLER. *Estimation of extreme expectiles from heavy tailed distributions*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics", Seville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01415586

[40] S. GIRARD, A. DAOUIA, G. STUPFLER. *Estimation of tail risk based on extreme expectiles*, in "Workshop Extremes - Copulas - Actuarial science", Luminy, France, February 2016, https://hal.archives-ouvertes.fr/hal-01311778

[41] S. GIRARD, A. DAOUIA, G. STUPFLER. *Tail risk estimation based on extreme Lp-quantiles*, in "Statistics workshop Tilburg University", Tilburg, Netherlands, December 2016, https://hal.archives-ouvertes.fr/hal-01415533

International Conferences with Proceedings

[42] M. ALBUGHDADI, L. CHAARI, F. FORBES, J.-Y. TOURNERET, P. CIUCIU. *Multi-subject joint parcellation detection estimation in functional MRI*, in "13th IEEE International Symposium on Biomedical Imaging", Prague, Czech Republic, April 2016, https://hal.inria.fr/hal-01261982

[43] P. FERNIQUE, A. DAMBREVILLE, J.-B. DURAND, C. PRADAL, P.-E. P.-E. LAURI, F. NORMAND, Y. GUÉDON. *Characterization of mango tree patchiness using a tree-segmentation/clusteringapproach*, in "2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA 2016)", Qingdao, China, November 2016, https://hal.inria.fr/hal-01398291

[44] B. PALLAS, W. YANG, J.-B. DURAND, S. S. MARTINEZ, E. E. COSTES. *Impact of Long Term Water Deficit on Production and Flowering Occurrence in the 'Granny Smith' Apple Tree Cultivar*, in "XI International Symposium on Integrating Canopy, Rootstock and Environmental Physiology in Orchard Systems", Bologna, Italy, XI International Symposium on Integrating Canopy, Rootstock and Environmental Physiology in Orchard Systems, Prof. Dr. Luca Corelli-Grappadelli, Department of Agricultural Sciences, Università di Bologna, August 2016, https://hal.inria.fr/hal-01377104

National Conferences with Proceedings

[45] J.-B. DURAND, A. GUÉRIN-DUGUÉ, S. ACHARD. *Analyse de séquences oculométriques et d'électroencéphalogrammes par modèles markoviens cachés*, in "48èmes Journées de Statistique", Montpellier, France, May 2016, https://hal.inria.fr/hal-01339458

Conferences without Proceedings

[46] C. ALBERT, A. DUTFOY, S. GIRARD. *Encadrement de l'erreur asymptotique d'estimation des quantiles extrêmes*, in "48èmes Journées de Statistique organisées par la Société Française de Statistique", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01326839

[47] J. ARBEL, I. PRÜNSTER. *Truncation error of a superposed gamma process in a decreasing order representation*, in "NIPS - 30th Conference on Neural Information Processing Systems", Barcelone, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01405580

[48] J. ARBEL, J.-B. SALOMOND. *Sequential Quasi Monte Carlo for Dirichlet Process Mixture Models*, in "NIPS - Conference on Neural Information Processing Systems", Barcelone, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01405568

[49] G. KON KAM KING, J. ARBEL, I. PRÜNSTER. *A Bayesian nonparametric approach to ecological risk assessment*, in "3rd Bayesian Young Statisticians Meeting (BAYSM 2016)", Florence, Italy, June 2016, https://hal.archives-ouvertes.fr/hal-01405593

[50] M. LOPES, M. FAUVEL, S. GIRARD, D. SHEEREN. *High dimensional Kullback-Leibler divergence for grassland management practices classification from high resolution satellite image time series*, in "IGARSS 2016 - IEEE International Geoscience and Remote Sensing Symposium", Bejing, China, July 2016, https://hal.archives-ouvertes.fr/hal-01366208

[51] M. LOPES, S. GIRARD, M. FAUVEL. *Divergence de Kullback-Leibler en grande dimension pour la classification des prairies à partir de séries temporelles d'images satellite à haute résolution*, in "48èmes Journées

de Statistique organisées par la Société Française de Statistique", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01326836

[52] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE. *Non linear robust regression in high dimension*, in "The XXVIIIth International Biometric Conference", Victoria, Canada, July 2016, https://hal.archives-ouvertes.fr/hal-01423622

[53] E. PERTHAME, F. FORBES, B. OLIVIER, A. DELEFORGE. *Regression non lineaire robuste en grande dimension*, in "48èmes Journées de Statistique organisées par la Société Française de Statistique", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01423630

Scientific Books (or Scientific Book chapters)

[54] F.-B. DIDIER, G. STÉPHANE (editors). *Statistics for Astrophysics: Clustering and Classification*, EAS Publications Series, EDP Sciences, Les Houches, France, 2016, vol. 77, https://hal.archives-ouvertes.fr/hal-01324665

[55] S. DOYLE, F. FORBES, M. DOJAT. *Automatic multiple sclerosis lesion segmentation with P-LOCUS*, in "Proceedings of the 1st MICCAI Challenge on Multiple Sclerosis Lesions Segmentation Challenge Using a Data Managementand Processing Infrastructure — MICCAI-MSSEG", 2016, pp. 17-21, http://www.hal.inserm.fr/inserm-01417434

[56] F. FORBES. *Modelling structured data with probabilistic graphical models*, in "Statistics for Astrophysics-Classification and Clustering", EDP Sciences, EAS Publication series, 2016, vol. 77, pp. 2016 - 2016, https://hal.archives-ouvertes.fr/hal-01423613

[57] S. GIRARD, J. SARACCO. *Supervised and unsupervised classification using mixture models*, in "Statistics for Astrophysics: Clustering and Classification", D. FRAIX-BURNET, S. GIRARD (editors), EAS Publications Series, EDP Sciences, May 2016, vol. 77, pp. 69-90, https://hal.archives-ouvertes.fr/hal-01417514

[58] C. MAGGIA, S. DOYLE, F. FORBES, O. HECK, I. TROPRÈS, C. BERTHET, Y. TEYSSIER, L. VELLY, J.-F. PAYEN, M. DOJAT. *Assessment of Tissue Injury in Severe Brain Trauma*, in "Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries", Lecture Notes in Computer Science, Springer International Publishing, 2016, vol. 9556, pp. 57-68, https://hal.archives-ouvertes.fr/hal-01423467

Other Publications

[59] M. ALBUGHDADI, L. CHAARI, J.-Y. TOURNERET, F. FORBES, P. CIUCIU. *Hemodynamic Brain Parcellation Using A Non-Parametric Bayesian Approach*, February 2016, working paper or preprint, https://hal.inria.fr/hal-01275622

[60] R. AZAÏS, J.-B. DURAND, C. GODIN. *Approximation of trees by self-nested trees*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01294013

[61] R. AZAÏS, J.-B. DURAND, C. GODIN. *Lossy compression of unordered rooted trees*, March 2016, DCC 2016 - Data Compression Conference , Poster [*DOI :* 10.1109/DCC.2016.73], https://hal.inria.fr/hal-01394707

[62] L. CHAARI, S. BADILLO, T. VINCENT, G. DEHAENE-LAMBERTZ, F. FORBES, P. CIUCIU. *Subject-level Joint Parcellation-Detection-Estimation in fMRI*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01255465

[63] A. DAOUIA, S. GIRARD, G. STUPFLER. *Estimation of Tail Risk based on Extreme Expectiles*, June 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01142130

[64] J. EL METHNI, L. GARDES, S. GIRARD. *Kernel estimation of extreme regression risk measures*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01393519

[65] P. FERNIQUE, J. LEGRAND, J.-B. DURAND, Y. GUÉDON. *Semi-parametric Markov Tree for cell lineage analysis*, June 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01286298

[66] P. FERNIQUE, J. PEYHARDI, J.-B. DURAND. *Multinomial distributions for the parametric modeling of multivariate count data*, April 2016, working paper or preprint, https://hal.inria.fr/hal-01286171

[67] F. FORBES. *Introduction to statistical methods in signal and image processing*, July 2016, Lecture, https://hal.archives-ouvertes.fr/cel-01423624

[68] M. LOPES, M. FAUVEL, S. GIRARD, D. SHEEREN, M. LANG. *High Dimensional Kullback-Leibler Divergence for grassland object-oriented classification from high resolution satellite image time series*, May 2016, Living Planet Symposium, Poster, https://hal.archives-ouvertes.fr/hal-01326865

[69] M. LOPES, M. FAUVEL, S. GIRARD, D. SHEEREN, M. LANG. *High Dimensional Kullback-Leibler Divergence for grassland object-oriented classification from high resolution satellite image time series*, March 2016, 4ème Journée Thématique du Programme National de Télédétection Spatiale (PNTS), Poster, https://hal.archives-ouvertes.fr/hal-01366221

[70] M. LOPES, M. M. FAUVEL, S. GIRARD, D. SHEEREN. *Object-based classification from high resolution satellite image time series with Gaussian mean map kernels: Application to grassland management practices*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01424929

[71] E. PERTHAME, F. FORBES, A. DELEFORGE. *Inverse regression approach to robust non-linear high-to-low dimensional mapping*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347455

[72] E. PERTHAME, C.-F. SHEU, D. CAUSEUR. *Signal identification in ERP data by decorrelated Higher Criticism Thresholding*, May 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01310739

References in notes

[73] S. BADILLO, T. VINCENT, P. CIUCIU. *Multi-session extension of the joint-detection framework in fMRI*, in "ISBI 2013 - International Symposium on BIomedical Imaging: From Nano to Macro", San Fransisco, United States, IEEE, April 2013, pp. 1512-1515 [*DOI :* 10.1109/ISBI.2013.6556822], https://hal.inria.fr/hal-00854624

[74] 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

[75] A. DELEFORGE, F. FORBES, R. HORAUD. *Acoustic Space Learning for Sound-Source Separation and Localization on Binaural Manifolds*, in "International Journal of Neural Systems", February 2015, vol. 25, n^{o} 1, 21p p. [*DOI :* 10.1142/S0129065714400036], https://hal.inria.fr/hal-00960796

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

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

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

[79] C. GODIN, P. FERRARO. *Quantifying the degree of self-nestedness of trees: application to the structural analysis of plants*, in "IEEE/ACM Transactions in Computational Biology and Bioinformatics", 2010, vol. 7, pp. 688–703, http://www-sop.inria.fr/virtualplants/Publications/2010/GF10

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

[81] P. MESEJO, S. SAILLET, O. DAVID, C. BÉNAR, J. M. WARNKING, F. FORBES. *Estimating Biophysical Parameters from BOLD Signals through Evolutionary-Based Optimization*, in "18th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'15)", Munich, Germany, October 2015, vol. Part II, pp. 528-535 [*DOI :* 10.1007/978-3-319-24571-3_63], https://hal.inria.fr/hal-01221126

[82] P. MESEJO, S. SAILLET, O. DAVID, C. BÉNAR, J. M. WARNKING, F. FORBES. *A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data*, in "IEEE Journal of Selected Topics in Signal Processing", March 2016, vol. 10, n^{o} 2, pp. 416-427 [*DOI :* 10.1109/JSTSP.2015.2502553], https://hal.inria.fr/hal-01221115

[83] R. NELSEN. *An introduction to copulas*, Lecture Notes in Statistics, Springer-Verlag, New-York, 1999, vol. 139

[84] F. SCHMIDT, J. FERNANDO. *Realistic uncertainties on Hapke model parameters from photometric measurements*, in "Icarus", 2015, vol. 260, pp. 73-93 (IF 2,84), https://hal.archives-ouvertes.fr/hal-01179842