



IN PARTNERSHIP WITH:  
**CNRS**

**Institut polytechnique de  
Grenoble**

**Université Joseph Fourier  
(Grenoble)**

## Activity Report 2015

# 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

<b>1. Members</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Research Program</b>	<b>3</b>
3.1. Mixture models	3
3.2. Markov models	3
3.3. Functional Inference, semi- and non-parametric methods	4
3.3.1. Modelling extremal events	4
3.3.2. Level sets estimation	6
3.3.3. Dimension reduction	6
<b>4. Application Domains</b>	<b>6</b>
4.1. Image Analysis	6
4.2. Biology, Environment and Medicine	6
<b>5. Highlights of the Year</b>	<b>7</b>
<b>6. New Software and Platforms</b>	<b>7</b>
6.1. MMST	7
6.2. P-LOCUS	7
6.3. PyHRF	8
<b>7. New Results</b>	<b>8</b>
7.1. Mixture models	8
7.2. Semi and non-parametric methods	8
7.2.1. Conditional extremal events	8
7.2.2. Estimation of extreme risk measures	9
7.2.3. Multivariate extremal events	9
7.2.4. Level sets estimation	10
7.2.5. Retrieval of Mars surface physical properties from OMEGA hyperspectral images.	10
7.2.6. Robust Sliced Inverse Regression.	11
7.2.7. Robust Locally linear mapping with mixtures of Student distributions	11
7.3. Markov models	11
7.3.1. Change-point models for tree-structured data	11
7.3.2. Hidden Markov models for the analysis of eye movements	11
7.3.3. Lossy compression of tree structures	12
7.4. Statistical models for Neuroscience	12
7.4.1. Comparison of stochastic and variational solutions to ASL fMRI data analysis	12
7.4.2. A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data	12
7.4.3. Multi-subject joint parcellation detection estimation in functional MRI	13
7.4.4. Tumor classification and prediction using robust multivariate clustering of multiparametric MRI	13
7.4.5. Functional specifications of a brain segmentation software	14
<b>8. Bilateral Contracts and Grants with Industry</b>	<b>14</b>
8.1. Bilateral Contracts with Industry	14
8.2. Bilateral Grants with Industry	14
<b>9. Partnerships and Cooperations</b>	<b>14</b>
9.1. Regional Initiatives	14
9.2. National Initiatives	15
9.2.1. Defi Imag'IN CNRS	15
9.2.2. GDR Madics	15
9.2.3. Networks	15
9.3. European Initiatives	15

---

9.4. International Initiatives	16
9.4.1. Inria International Labs	16
9.4.2. Inria International Partners	16
9.5. International Research Visitors	17
9.5.1.1. Researchers	17
9.5.1.2. Internships	17
<b>10. Dissemination</b> .....	<b>17</b>
10.1. Promoting Scientific Activities	17
10.1.1. Scientific events organisation	17
10.1.2. Scientific events selection	17
10.1.2.1. Member of the conference program committees	17
10.1.2.2. Reviewer	17
10.1.3. Journal	18
10.1.3.1. Member of the editorial boards	18
10.1.3.2. Reviewer - Reviewing activities	18
10.1.4. Invited talks	18
10.1.5. Leadership within the scientific community	18
10.1.6. Scientific expertise	18
10.1.7. Research administration	18
10.2. Teaching - Supervision - Juries	19
10.2.1. Teaching	19
10.2.2. Supervision	19
10.2.3. Juries	19
10.2.3.1. PhD	19
10.2.3.2. HDR	20
10.2.3.3. Other committees	20
10.3. Popularization	20
<b>11. Bibliography</b> .....	<b>20</b>

# 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 [Inria, Team leader, Senior Researcher, HdR]  
Stephane Girard [Inria, Senior Researcher, since Oct 2015, HdR]

### Faculty Members

Jean Baptiste Durand [INP Grenoble, Associate Professor]  
Marie-Jose Martinez [Univ. Grenoble II, Associate Professor, until May 2015]

### Engineers

Thomas Perret [Inria]  
Pascal Rubini [Inria, from Apr 2015]

### PhD Students

Clément Albert [Inria, from Mar 2015]  
Alexis Arnaud [Univ. Grenoble I]  
Alessandro Chiancone [Univ. Grenoble I]  
Aina Frau Pascual [Inria]  
Brice Olivier [Univ. Grenoble I, from Mar 2015]  
Thibaud Rahier [Schneider Electric, from Oct 2015]  
Pierre-Antoine Rodesch [CEA, from Oct 2015]

### Post-Doctoral Fellows

Pablo Mesejo Santiago [Inria]  
Emeline Perthame [Inria, from Nov 2015]

**Visiting Scientists**

Mohanad Albughdadi [IRIT, from Feb 2015 until Mar 2015]  
El Hadji Deme [Univ. Gaston Berger, Senegal, from Oct 2015 until Nov 2015]  
Mailys Lopes [INRA, from Nov 2015]  
Abdelhakim Necir [Univ Biskra, Algeria, from Oct 2015 until Nov 2015]

**Administrative Assistants**

Giselle Loquet [Inria, until Jun 2015]  
Myriam Etienne [Inria, from July 2015]

**Others**

Léo Nicoletti [Ensimag, Intern, from Feb 2015 until May 2015]  
Priscillia Previtero [ENSBM, Intern, from Mar 2015 until Aug 2015]  
Seydou Nourou Sylla [Univ. Gaston Berger, Senegal, PhD]  
Sebastian Torres Leiva [Univ Santiago, Chile, Intern, from Feb 2015 until Jun 2015]

## 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, these 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 connected with 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, Stéphane Girard, Marie-José Martinez.

**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,  $\theta$  being the parameter, possibly multi-dimensional, usually unknown and to be estimated. We consider cases where the data naturally divides into observed data  $y = y_1, \dots, y_n$  and unobserved or missing data  $z = z_1, \dots, z_n$ . 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,

$$f(y_i | \theta) = \sum_{k=1}^K P(z_i = k | \theta) f(y_i | z_i, \theta). \quad (1)$$

These models are interesting in that they may point out hidden variable 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.

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

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

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

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

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

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

**Participants:** Clément Albert, Alessandro Chiancone, Stéphane Girard, Seydou-Nourou Sylla, Pablo Mesejo, Florence Forbes.

**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 [63]. 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 [66] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [62], 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  $X_{1,n} \leq \dots \leq X_{n,n}$  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).



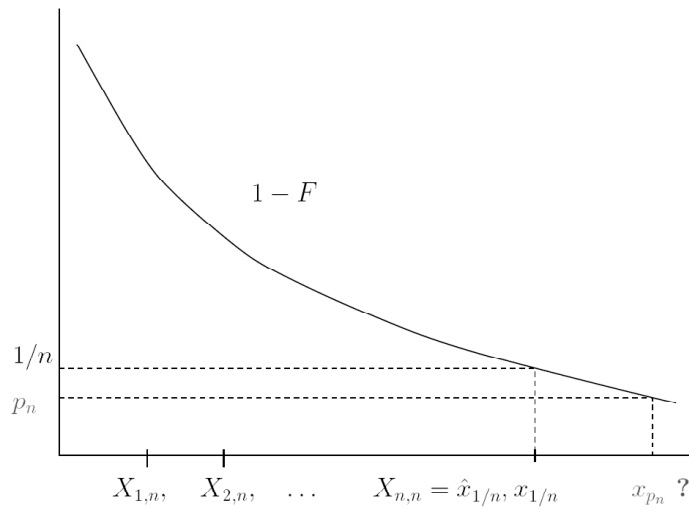


Figure 1. The curve represents the survival function  $x \rightarrow P(X > x)$ . The  $1/n$ -quantile is estimated by the maximum observation so that  $\hat{x}_{1/n} = X_{n,n}$ . As illustrated in the figure, to estimate  $p_n$ -quantiles with  $p_n < 1/n$ , it is necessary to extrapolate beyond the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

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

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

$$\frac{\ell(tx)}{\ell(x)} \rightarrow 1 \quad \text{as } x \rightarrow \infty \quad (3)$$

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  $\rho$  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:

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

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

### 3.3.2. Level sets estimation

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

### 3.3.3. Dimension reduction

Our work on high dimensional data requires that we face the curse of dimensionality phenomenon. Indeed, the modelling of high dimensional data requires complex models and thus the estimation of high number of parameters compared to the sample size. In this framework, dimension reduction methods aim at replacing the original variables by a small number of linear combinations with as small as a possible loss of information. Principal Component Analysis (PCA) is the most widely used method to reduce dimension in data. However, standard linear PCA can be quite inefficient on image data where even simple image distortions can lead to highly non-linear data. Two directions are investigated. First, non-linear PCAs can be proposed, leading to semi-parametric dimension reduction methods [64]. 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 [61]. 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 [66].

## 4. Application Domains

### 4.1. Image Analysis

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

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 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 competitivite" project I-VP, we worked of images of PC Boards.

### 4.2. Biology, Environment and Medicine

**Participants:** Pablo Mesejo Santiago, Aina Frau Pascual, Florence Forbes, Stéphane Girard, Seydou-Nourou Sylla, Marie-José Martinez, Jean-Baptiste Durand.

A second domain of applications concerns biology and medicine. We consider 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. Finally, in the context of the ANR VMC project Medup, we studied the uncertainties on the forecasting and climate projection for Mediterranean high-impact weather events.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

- Creation of the Pixyl startup (<http://pixyl.io>). Pixyl is a startup created in March 2015 by F. Forbes with M. Dojat (INSERM), a former post-doctoral fellow S. Doyle (CEO) and IT Translation. F. Forbes is a co-founder and a scientific advisor. Pixyl specializes in automatic pathological brain MRI segmentation. Pixyl delivers precise neuroimaging biomarker extraction for improved insight and decision-making in clinical studies. Our automatic neuroimaging solutions are designed for seamless integration into both web-based and locally deployed CRO platforms. Advanced algorithms are available for multiple sclerosis, chronic stroke, brain tumor and traumatic brain injury applications. The technology transferred is based on the P-LOCUS software.
- Xerox Foundation University Affairs Committee (UAC) collaborative grant. F. Forbes was co-laureate (with R. Horaud) of this grant (90 k\$) for a three year project (2014-2017) on Advanced and Scalable Graph Signal Processing Techniques. It was awarded in collaboration with Arijit Biswas and Anirban Mondal, research scientists at Xerox Research Center India (XRCI), Bangalore.

## 6. New Software and Platforms

### 6.1. 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 generalisation of multivariate Student T distribution allowing different tails in each dimension. Typical applications include Robust clustering to analyse data with possible outliers. In this context, the model and package have been used on large data sets of brain MRI to segment and identify brain tumors.

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

### 6.2. P-LOCUS

KEYWORDS: Health - Neuroimaging - Cancer - Brain MRI - Medical imaging

FUNCTIONAL DESCRIPTION

The Locus software was extended to address the delineation of lesions in pathological brains. Its extension P-LOCUS software analyses, in few minutes, a 3D MR brain scan and performs fully automatic brain lesion delineation using a combined dataset of various 3D MRI sequences.

- Participants: Senan Doyle, Florence Forbes, Michel Dojat and Pascal Rubini
- Partner: INSERM
- Contact: Florence Forbes
- URL: <http://p-locus.com/>

## 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 and Aina Frau Pascual
- Partners: CEA - NeuroSpin
- Contact: Florence Forbes
- URL: <http://pyhrf.org>

## 7. New Results

### 7.1. Mixture models

#### 7.1.1. Taking into account the curse of dimensionality

**Participants:** Stéphane Girard, Alessandro Chiancone, Seydou-Nourou Sylla.

**Joint work with:** C. Bouveyron (Univ. Paris 5), M. Fauvel (ENSAT Toulouse) and J. Chanussot (Gipsa-lab and Grenoble-INP)

In the PhD work of Charles Bouveyron (co-advised by Cordelia Schmid from the Inria LEAR team) [61], 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 yields 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]. Our recent work consists in adding a kernel in the previous methods to deal with nonlinear data classification and heterogeneous data [13]. We first investigate the use of kernels derived from similiary measures on binary data [30]. The targeted application is the analysis of verbal autopsy data (PhD thesis of N. Sylla): Indeed, health monitoring and evaluation make more and more use of data on causes of death from verbal autopsies in countries which do not keep records of civil status or with incomplete records. The application of verbal autopsy method allows to discover probable cause of death. Verbal autopsy has become the main source of information on causes of death in these populations. Second, the kernel classification method is applied to three real hyperspectral data sets, and compared with three others classifiers. The proposed models show good results in terms of classification accuracy and processing time [21].

### 7.2. Semi and non-parametric methods

#### 7.2.1. Conditional extremal events

**Participant:** Stéphane Girard.

**Joint work with:** L. Gardes (Univ. Strasbourg), G. Mazo (Univ. Catholique de Louvain), J. Elmethni (Univ. Paris 5) and S. Louhichi (Univ. Grenoble 1)

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 [63] with extreme-value methods in order to obtain efficient estimators of the conditional tail index and conditional extreme quantiles. The strong consistency of such an estimator is established in [53]. When the covariate is functional and random (random design) we focus on kernel methods [23].

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 [20] and the "département Génie urbain" of "Université Paris-Est Marne-la-vallée" [11].

### 7.2.2. *Estimation of extreme risk measures*

**Participant:** Stéphane Girard.

**Joint work with:** A. Daouia (Univ. Toulouse), E. Deme (Univ. Gaston-Berger, Sénégal), A. Guillou (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  $\alpha$ -quantile of the loss distribution. The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information:  $\text{VaR}(\alpha)$  does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper  $\alpha$ -quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution. It is frequently encountered in financial investment or in the insurance industry. In [51], we have established the asymptotic properties of the CTE estimator in case of extreme losses, *i.e.* when  $\alpha \rightarrow 0$  as the sample size increases. We have exhibited the asymptotic bias of this estimator, and proposed a bias correction based on extreme-value techniques. A second possible coherent alternative risk measure is based on expectiles [59]. 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.

### 7.2.3. *Multivariate extremal events*

**Participants:** Stéphane Girard, Florence Forbes.

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

Copulas are a useful tool to model multivariate distributions [67].

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. We proposed a new class of high dimensional copulas based on a product of transformed bivariate copulas [26]. No constraints on the parameters refrain the applicability of the proposed class which is well suited for applications in high dimension. Furthermore the analytic forms of the copulas within this class allow to associate a natural graphical structure which helps to visualize the dependencies and to compute the likelihood efficiently even in high dimension. The extreme properties of the copulas are also derived and an R package has been developed.

As an alternative, we also 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 [19], [27], [52]. A pairwise moment-based inference procedure has also been proposed and the asymptotic normality of the corresponding estimator has been established [28].

In collaboration with L. Gardes, we investigate the estimation of the tail copula, which is widely used to describe the amount of extremal dependence of a multivariate distribution. In some situations such as risk management, the dependence structure can be linked with some covariate. The tail copula thus depends on this covariate and is referred to as the conditional tail copula. The aim of our work is to propose a nonparametric estimator of the conditional tail copula and to establish its asymptotic normality [22].

#### 7.2.4. Level sets estimation

**Participant:** Stéphane 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 [24], [25] and thus should not be used to detect outliers.

#### 7.2.5. Retrieval of Mars surface physical properties from OMEGA hyperspectral images.

**Participants:** Stéphane Girard, Alessandro Chiancone.

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

Visible and near infrared imaging spectroscopy is one of the key techniques to detect, to map and to characterize mineral and volatile (eg. water-ice) species existing at the surface of planets. Indeed the chemical composition, granularity, texture, physical state, etc. of the materials determine the existence and morphology of the absorption bands. The resulting spectra contain therefore very useful information. Current imaging spectrometers provide data organized as three dimensional hyperspectral images: two spatial dimensions and one spectral dimension. Our goal is to estimate the functional relationship  $F$  between some observed spectra and some physical parameters. To this end, a database of synthetic spectra is generated by a physical radiative transfer model and used to estimate  $F$ . The high dimension of spectra is reduced by Gaussian regularized sliced inverse regression (GRSIR) to overcome the curse of dimensionality and consequently the sensitivity of the inversion to noise.

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 [14].

### 7.2.6. Robust Sliced Inverse Regression.

**Participants:** Stéphane Girard, Alessandro Chiancone, Florence Forbes.

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 been proposed to robustify SIR, in this work, 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 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. Robust Locally linear mapping with mixtures of Student distributions

**Participants:** Florence Forbes, Emeline Perthame, Brice Olivier, Leo Nicoletti.

The standard GLLiM model [17] 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.

## 7.3. Markov models

### 7.3.1. 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 [45].

### 7.3.2. 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 through eye-movements. Our approach was inspired by Simola *et al.* (2008) [68], but we used hidden semi-Markov models for better characterization of phase length distributions. The estimated HMM highlighted contrasted reading strategies (i.e., 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.

### 7.3.3. *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 [65], 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 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 in a conference paper accepted at DCC'2016 [55] with 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 [65] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. In addition, we obtained new theoretical results on the combinatorics of self-nested structures. The redaction of an article is currently in progress.

## 7.4. Statistical models for Neuroscience

### 7.4.1. *Comparison of stochastic and variational solutions to ASL fMRI data analysis*

**Participants:** Florence Forbes, Aina Frau Pascual.

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

Functional Arterial Spin Labeling (fASL) MRI can provide a quantitative measurement of changes of cerebral blood flow induced by stimulation or task performance. fASL data is commonly analysed using a general linear model (GLM) with regressors based on the canonical hemodynamic response function. In this work [37], we consider instead a joint detection-estimation (JDE) framework which has the advantage of allowing the extraction of both task-related perfusion and hemodynamic responses not restricted to canonical shapes. Previous JDE attempts for ASL have been based on computer intensive sampling (MCMC) methods. Our contribution is to provide a comparison with an alternative variational expectation-maximization (VEM) algorithm on synthetic and real data. Other investigations were related to the use of appropriate physiological information and priors [39], [38].

### 7.4.2. *A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data*

**Participants:** Florence Forbes, Pablo Mesejo.



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

Physiological and biophysical models have been proposed to link neuronal activity to the Blood Oxygen Level-Dependent (BOLD) signal in functional MRI (fMRI). Those models rely on a set of parameter values that cannot always be extracted from the literature. In some applications, interesting insight into the brain physiology or physiopathology can be gained from an estimation of the model parameters from measured BOLD signals. This estimation is challenging because there are more than 10 potentially interesting parameters involved in nonlinear equations and whose interactions may result in identifiability issues. However, the availability of statistical prior knowledge about these parameters can greatly simplify the estimation task. In this work we focus on the extended Balloon model and propose the estimation of 15 parameters using two stochastic approaches: an Evolutionary Computation global search method called Differential Evolution (DE) and a Markov Chain Monte Carlo version of DE. To combine both the ability to escape local optima and to incorporate prior knowledge, we derive the target function from Bayesian modeling. The general behavior of these algorithms is analyzed and compared with the *de facto* standard Expectation Maximization Gauss-Newton (EM/GN) approach, providing very promising results on challenging real and synthetic fMRI data sets involving rats with epileptic activity. These stochastic optimizers provided a better performance than EM/GN in terms of distance to the ground truth in 4 out of 6 synthetic data sets and a better signal fitting in 12 out of 12 real data sets. Non-parametric statistical tests showed the existence of statistically significant differences between the real data results obtained by DE and EM/GN. Finally, the estimates obtained from DE for these parameters seem both more realistic and more stable or at least as stable across sessions as the estimates from EM/GN. This work will appear in [29]. A preliminary version has also been accepted at the conference MICCAI 2015 [40].

#### 7.4.3. *Multi-subject joint parcellation detection estimation in functional MRI*

**Participant:** Florence Forbes.

**Joint work with:** Lotfi Chaari, Mohanad Albughdadi, Jean-Yves Tournet from IRIT-ENSEEIH 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 accepted at ISBI 2016.

#### 7.4.4. *Tumor classification and prediction 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.

Advanced statistical clustering approaches are promising tools to better exploit the wealth of MRI information especially on large cohorts and multi-center studies. In neuro-oncology, the use of multiparametric MRI may better characterize brain tumor heterogeneity. To fully exploit multiparametric MRI (e.g. tumor classification), appropriate analysis methods are yet to be developed. They offer improved data quality control by allowing automatic outlier detection and improved analysis by identifying discriminative tumor signatures with measurable predictive power. In this work, we show on small animals data that advanced statistical learning approaches can help 1) in organizing existing data by detecting and excluding outliers and 2) in building a dictionary of tumor fingerprints from a clustering analysis of their microvascular features. The work also now

includes the integration in a joint statistical model of both automatic ROI delineation and clustering for whole brain data analysis. A preliminary version of this work has been accepted to the ISMRM 2015 conference and in the SFMRMB 2015 conference [41].

#### 7.4.5. *Functional specifications of a brain segmentation software*

**Participants:** Florence Forbes, Priscillia Previtero.

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

The goal of P. Previtero's internship was to help with a number of software engineering tasks and communications actions around the P-Locus software and the Pixyl start-up. The internship resulted in particular in a new web site for Pixyl.

## 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

**DGARapid WIFUZ (2015-2017).** F. Forbes is the principal investigator for MISTIS of the 2 year project WIFUZ on *W*ireless multi sensors *F*USion. The project is supported by DGA and led by the ACOEM company <http://www.acoemgroup.fr/> and involves another company, the HIKOB startup, <http://www.hikob.com/>. The objective is to develop a prototype for surveillance and monitoring that is able to combine multi sensor data coming from acoustic sensors (microphones or antennas) and optical sensors (infrared cameras) and to distribute the processing to multiple algorithmic blocs. The financial support for MISTIS is of 122,4 keuros for a project of a total cost of 375 keuros.

**CIFRE PhD with SCHNEIDER (2015-2018).** F. Forbes and S. Girard are the advisors of a starting 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 will be of 165 keuros.

**PhD contract with EDF (2015-2018).** S. Girard is the advisor of a starting PhD (A. Clement) with EDF. The goal is to investigate sensitivity analysis and extrapolation limits in Extreme value theory with application to river flows analysis.

### 8.2. Bilateral Grants with Industry

**UAC XEROX INDIA (2014-2017).** F. Forbes is co-principal investigator with R. Horaud (PERCEPTION) of a Xerox Foundation University Affairs Committee (UAC) collaborative grant *Advanced and Scalable Graph Signal Processing Techniques*, in collaboration with Arijit Biswas and Anirban Mondal, research scientists at Xerox Research Center India (XRCI) Bangalore. This collaboration is an opportunity to launch a joint research program with a Xerox Indian team. We plan to investigate robust mixture models and techniques to deal with graphical data. Xerox Foundation funding: 80 keuros.

## 9. Partnerships and Cooperations

### 9.1. Regional Initiatives

- **PERSYVACT projects.**

MISTIS is involved in the 3-year project-team Oculo Nimbus, funded (250 keuros for the whole project) by the PERSYVAL labex (<https://persyval-lab.org/en>), with other teams from local laboratories, LJK, GIPSA-Lab and LPNC. The goal of this research project is to develop tools for analyzing eye-movement data.

MISTIS is also involved in another action (2015-2018) recently granted Persyvact2 action supported by the Persyval Labex for 3.5 years. This project is a follow-up of the Persyvact Exploratory labex project. Persyvact2 consists of about 20 researchers from different laboratories, GIPSA-lab, LJK and TIMC-IMAG and different fields related to data science (statistics, machine learning, image and signal processing). Our contribution and involvement will lie essentially in a Graph signal processing work package with application in neuroscience for which we are planning to hire a PhD student with S. Achard (GIPSA-Lab). Persyvact2 also intends to organize scientific events and an international workshop during its lifetime. Persyvact2 will contribute, with other teams of Persyval, to enhance the international visibility of data science in Grenoble. The financial support for the consortium is of 250 keuros.

- **Grenoble Pole Cognition (2013-15).** We received in 2015 2.5 keuros from the Grenoble Pole Cognition, <http://www.grenoblecognition.fr/>, for collaborative projects involving the GIN and the Pixyl startup. This funding was used this year for the internship of Priscillia Previtero on brain MRI analysis.
- 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.

## 9.2. National Initiatives

### 9.2.1. 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 made in the HUMAVIPS project. The CNRS support for the network is of 20 keuros.

### 9.2.2. 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).

### 9.2.3. 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. European Initiatives

### 9.3.1. FP7 & H2020 Projects

**European H2020 RESSTORE (2015-2018).** F. Forbes is involved in this multi-center Stroke European H2020 project including 20 partners. F. Forbes will contribute through the Pixyl startup which will receive 70 keuros as a subcontractor. RESSTORE stands for REgenerative Stem cell therapy for STroke in Europe. It is part of the Clinical research on regenerative medicine program. It will involve a phase 2 trial with 300 patients imaged at 4 time points over a 3 year timeframe. Pixyl will provide automatic stroke lesion segmentations.

## 9.4. International Initiatives

### 9.4.1. Inria International Labs

#### LIRIMA

Associate Team involved in the International Lab:

##### 9.4.1.1. SIMERGE

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

International Partner (Institution - Laboratory - Researcher):

UGB (Senegal) - LERSTAD - Abdou Kâ Diongue

Start 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.

### 9.4.2. Inria International Partners

#### 9.4.2.1. Informal International Partners

The context of our research is also the collaboration between MISTIS and a number of international partners such as the Statistics Department of University of Washington in Seattle, the Russian Academy of Science in Moscow, and more recent partners like IDIAP involved in the past HUMAVIPS project, Université Gaston Berger in Senegal and Universities of Melbourne and Brisbane in Australia. We also work at turning other current European contacts, *e.g.* at EPFL (A. Roche at University Hospital Lausanne and Siemens Healthcare), into more formal partnerships.

The main international collaborations that we are currently trying to develop are with:

- Fabrizio Durante, Free University of Bozen-Bolzano, Italy.
- K. Qin and D. Wraith 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.
- Alexandre Nazin and Russian Academy of Science in Moscow, Russia.
- Alexis Roche and University Hospital Lausanne/Siemens Healthcare, Advanced Clinical Imaging Technology group, Lausanne, Switzerland.

## 9.5. International Research Visitors

### 9.5.1. Visits of International Scientists

#### 9.5.1.1. Researchers

- Seydou Nourou Sylla (Université Gaston Berger, Sénégal) has been hosted by the MISTIS team for four months.
- El Hadji Deme has been hosted by the MISTIS team for 3 weeks.
- Abdelhakim Necir (University Biskra, Algeria) has been hosted for 2 weeks.

#### 9.5.1.2. Internships

Sebastian Torres Leiva (Master, from Feb 2015 until June 2015)

Subject: Extreme value modelling of some glacial processes in Chilean Andes.

Institution: UTFSM - Universidad Tecnica Federico Santa Maria, Valparaiso, Chile

## 10. Dissemination

### 10.1. Promoting Scientific Activities

#### 10.1.1. Scientific events organisation

##### 10.1.1.1. Member of the organizing committees

- **IEEE DSAA 2015, IEEE international conference on Data Science and Advanced Analytics**, <http://dsaa2015.lip6.fr/>, in Paris, Oct. 19-21, 2015. F. Forbes was member of the scientific committee and organized with W. Buntine a special session on "**Trends and controversies**". 150 participants. 81 accepted papers. Acceptance ratio 29%.
- Stéphane Girard was co-chair of the Astrostatistics summer school dedicated to "Classification & Clustering" held in Les Houches, <http://stat4astro2015.sciencesconf.org/>.
- **2nd conference of the SFRMBM society (Société Française de Résonance Magnétique en Biologie et Médecine)**, <http://sfrmbm2015.sciencesconf.org/>, March 18-20, 2015, Grenoble. F. Forbes was a member of the organizing committee. 200 participants. 140 submissions. Approximate funding: 80 keuros.
- Stéphane Girard was a member of the organizing committees of "4èmes rencontres R" <http://r2015-grenoble.sciencesconf.org/> and of the 6th workshop Statlearn "Challenging problems in Statistical Learning", both held in Grenoble.
- **INRA AIGM network day** in Grenoble June 30, 2015. F. Forbes was program co-chair with N. Peyrard. 20 participants. Approximate funding: 1 keuros. AIGM events web site: <https://carlit.toulouse.inra.fr/AIGM/events.html>.

#### 10.1.2. Scientific events selection

##### 10.1.2.1. Member of the conference program committees

- 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.

##### 10.1.2.2. Reviewer

In 2015, F. Forbes has been reviewer for NIPS 2015 and GRETSI 2015.

### 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. 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 2015, S. Girard has been a reviewer for *Scandinavian Journal of Statistics*, *Extremes* and *Journal of Statistical Software*.

In 2015, F. Forbes has been reviewer for *Statistics and Computing*, *Computational Statistics and Data Analysis* journals, *IEEE trans. on Signal Processing*, *IEEE trans. on Image Processing* journals.

### 10.1.4. Invited talks

Stéphane Girard was invited to give a talk at the SMAI Conference [32] and at the Extreme Value Analysis conference [33].

F. Forbes was invited to give a talk at the Working Group on Model-Based Clustering Summer Session, in Seattle, USA, July 19-25, 2015, <http://mathsci.ucd.ie/~brendan/wgmbc2015.html>. Title: High dimensional regression with Gaussian mixtures and partially latent response variables. 30 participants.

F. Forbes was invited to give a talk at the CHUV/SIEMENS workshop on Quantitative magnetic resonance imaging for neuroradiology, in Lausanne, Switzerland, June 29, 2015. Title: Automatic brain lesion segmentation: methodological challenges. 20 participants.

F. Forbes and S. Girard gave a tutorial at the Astrostatistics School in Les Houches in Oct. 2015, <http://stat4astro2015.sciencesconf.org/>.

### 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 Mistis and LERSTAD (Université Gaston Berger, Saint-Louis, Sénégal). The team is part of the LIRIMA (Laboratoire International de Recherche en Informatique et Mathématiques Appliquées), <http://mistis.inrialpes.fr/simerge/>.

### 10.1.6. Scientific expertise

- Stéphane Girard was in charge of evaluating PEPS research projects (projets exploratoires premier soutien) for the CNRS and MITACS projects from Québec, Canada.
- F. Forbes is a member of the ERCIM working group on Mixture models.

### 10.1.7. Research administration

- Stéphane Girard is at the head of the Probability and Statistics department of the LJK (Laboratoire Jean Kuntzmann) since september 2012.
- 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.
- F. Forbes was elected in 2010 and is since then a member of the bureau of the “Statistics and Images” group in the Société Française de Statistique (SFdS), [http://www.sfds.asso.fr/318-Actualites\\_du\\_groupe\\_Statistique\\_et\\_Images](http://www.sfds.asso.fr/318-Actualites_du_groupe_Statistique_et_Images). The group is made of 5 persons. F. Forbes is vice-president.

## 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching

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

Master : Stéphane Girard, *Statistique Inférentielle Avancée*, 45 ETD, M1 level, Ensimag. Grenoble-INP, France.

Master: Jean-Baptiste Durand, *Statistics and probability*, 192 ETD, M1 and M2 levels, Ensimag Grenoble INP, France.

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

### 10.2.2. Supervision

PhD in progress: Aina Frau-Pascual, “*Statistical Models for the coupling of ASL and BOLD Magnetic Resonance modalities to study brain function and disease*”, October 2013, Florence Forbes and Philippe Ciuciu.

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.

PhD in progress: Pierre-Antoine Rodesch, “*Spectral tomography and tomographic reconstruction algorithms*”, october 2015, Florence Forbes.

PhD in progress: Thibaud Rahier, “*Data-mining pour la fusion de données structurées et non-structurées*”, november 2015, Florence Forbes and Stéphane Girard.

PhD in progress: Clément Albert, “*Limites de crédibilité d’extrapolation des lois de valeurs extrêmes*”, october 2015, Stéphane Girard.

PhD in progress: Maïlys Lopes, “*Téledétection en écologie du paysage : statistiques en grande dimension pour la multirésolution spatiale et la haute résolution temporelle*”, november 2014, Stéphane Girard and Mathieu Fauvel (INRA Toulouse).

PhD in progress: Alessandro Chiancone, “*Sequential dimension reduction*”, november 2013, Stéphane Girard and Jocelyn Chanussot (Grenoble INP).

PhD in progress: Seydou Nourou Sylla, “*Modélisation statistique pour l’analyse des causes de décès décrites par autopsie verbale en milieu rural africain : cas du Sénégal*”, october 2012, Stéphane Girard and Abdou Diongue (Université Gaston Berger, Sénégal).

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

- F. Forbes has been reviewer for four PhD theses in 2015:
  - Theodosios Gkamas, University of Strasbourg, Sept. 29, 2015.
  - Brice Ozenne, University of Lyon 1, Sept, 2015.
  - Julien Stoehr, University Montpellier 2, Oct. 2015.
  - Hajer Braham, Telecom ParisTech and Orange, December 2015.
- F. Forbes was also president for the PhD committee of Haithem Boussaid, "Efficient Inference and learning in Graphical models for multi-organ shape segmentation", Ecole Centrale Paris, January 8, 2015.



- Stéphane Girard has been reviewer of two PhD theses:
  - “*Modélisation de la dépendance et estimation du risque agrégé*”, by Andrés Cuberos, Univ. Claude Bernard - Lyon and
  - “*Analyse de données de cytométrie de flux pour un grand nombre d’échantillons*” by Xiaoyi Chen, Univ. Cergy-Pontoise.
- S. Girard was also a member of the PhD committee of Jonathan Jalbert “*Développement d’un modèle statistique non stationnaire et régional pour les précipitations extrêmes simulées par un modèle numérique de climat*”, Univ. Laval (Québec, Canada).

#### 10.2.3.2. HDR

F. Forbes was also a member of the HDR committee of Estelle Kuhn, University Paris Orsay, November 2015.

#### 10.2.3.3. Other committees

- F. Forbes is a member of the Committee for technological project and engineer candidate selection at Inria Grenoble Rhône-Alpes ("Commission du développement technologique").
- F. Forbes was a member of the committee for Inria research scientist candidate (CR) selection at Inria Rennes in 2015.
- F. Forbes was a member of the committee for attributing the bi-annual Jean Kuntzman award.
- Stéphane Girard is a member of the "Comité des Emplois Scientifiques" and “Comité de Centre” at Inria Grenoble Rhône-Alpes since 2015.
- Since 2015, Stéphane 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.

## 10.3. Popularization

Pixyl communication. F. Forbes gave several large audience presentations related to the Pixyl startup and more recently interviews for journals such as **Presence** and INSERM Science et Santé Magazines.

# 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<sup>o</sup> 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 Analysis and Machine Intelligence", 2008, vol. 30(6), pp. 1055–1067
- [3] C. BOUYEYRON, S. GIRARD, C. SCHMID. *High dimensional data clustering*, in "Computational Statistics and Data Analysis", 2007, vol. 52, pp. 502–519
- [4] C. BOUYEYRON, S. GIRARD, C. SCHMID. *High dimensional discriminant analysis*, in "Communication in Statistics - Theory and Methods", 2007, vol. 36, n<sup>o</sup> 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<sup>o</sup> 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<sup>o</sup> 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<sup>o</sup> 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<sup>o</sup> 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<sup>o</sup> 2, pp. 369-384

## Publications of the year

### Articles in International Peer-Reviewed Journals

- [11] B. BARROCA, P. BERNADARA, S. GIRARD, G. MAZO. *Considering hazard estimation uncertain in urban resilience strategies*, in "Natural Hazards and Earth System Sciences", 2015, vol. 15, pp. 25-34 [DOI : 10.5194/nhess-15-25-2015], <https://hal.archives-ouvertes.fr/hal-01100539>
- [12] C. BAZZOLI, F. LETUÉ, M.-J. MARTINEZ. *Modelling finger force produced from different tasks using linear mixed models with lme R function*, in "Journal of Case Studies in Business, Industry and Government Statistics (CSBIGS)", 2015, vol. 6, n<sup>o</sup> 1, pp. 16-36, <https://hal.archives-ouvertes.fr/hal-00998910>
- [13] C. BOUVEYRON, M. FAUVEL, S. GIRARD. *Kernel discriminant analysis and clustering with parsimonious Gaussian process models*, in "Statistics and Computing", 2015, vol. 25, n<sup>o</sup> 6, pp. 1143-1162, 33 pages - arXiv:1204.4021 [DOI : 10.1007/s11222-014-9505-x], <https://hal.archives-ouvertes.fr/hal-00687304>
- [14] A. CHIANCONE, S. GIRARD, J. CHANUSSOT. *Collaborative Sliced Inverse Regression*, in "Communication in Statistics - Theory and Methods", 2015, forthcoming, <https://hal.inria.fr/hal-01158061>
- [15] A. DELEFORGE, F. FORBES, S. BA, R. HORAUD. *Hyper-Spectral Image Analysis with Partially-Latent Regression and Spatial Markov Dependencies*, in "IEEE Journal on Selected Topics in Signal Processing", September 2015, vol. 9, n<sup>o</sup> 6, pp. 1037-1048 [DOI : 10.1109/JSTSP.2015.2416677], <https://hal.inria.fr/hal-01136465>
- [16] 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<sup>o</sup> 1, 21 p. [DOI : 10.1142/S0129065714400036], <https://hal.inria.fr/hal-00960796>
- [17] A. DELEFORGE, F. FORBES, R. HORAUD. *High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables*, in "Statistics and Computing", September 2015, vol. 25, n<sup>o</sup> 5, pp. 893-911 [DOI : 10.1007/s11222-014-9461-5], <https://hal.inria.fr/hal-00863468>

- [18] 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<sup>o</sup> 1, pp. 549-567 [DOI : 10.1007/s11222-014-9494-9], <https://hal.inria.fr/hal-01090836>
- [19] F. DURANTE, S. GIRARD, G. MAZO. *Marshall–Olkin type copulas generated by a global shock*, in "Journal of Applied and Computational Mathematics", 2016, vol. 296, pp. 638–648, <https://hal.archives-ouvertes.fr/hal-01138228>
- [20] J. EL METHNI, L. GARDES, S. GIRARD. *Estimation of risk measures for extreme pluviometrical measurements in the Cévennes-Vivarais region*, in "La Houille Blanche", September 2015, n<sup>o</sup> 4, pp. 46-51, <https://hal.archives-ouvertes.fr/hal-01212105>
- [21] M. FAUVEL, C. BOUVEYRON, S. GIRARD. *Parsimonious Gaussian process models for the classification of hyperspectral remote sensing images*, in "IEEE Geosciences and Remote Sensing Letters", 2015, vol. 12, pp. 2423–2427 [DOI : 10.1109/LGRS.2015.2481321], <https://hal.archives-ouvertes.fr/hal-01203269>
- [22] L. GARDES, S. GIRARD. *Nonparametric estimation of the conditional tail copula*, in "Journal of Multivariate Analysis", 2015, vol. 137, pp. 1–16, <https://hal.archives-ouvertes.fr/hal-00964514>
- [23] L. GARDES, S. GIRARD. *On the estimation of the functional Weibull tail-coefficient*, in "Journal of Multivariate Analysis", 2015, forthcoming, <https://hal.archives-ouvertes.fr/hal-01063569>
- [24] S. GIRARD, G. STUPFLER. *Extreme geometric quantiles in a multivariate regular variation framework*, in "Extremes", 2015, vol. 18, n<sup>o</sup> 4, pp. 629–663, <https://hal.archives-ouvertes.fr/hal-01155112>
- [25] S. GIRARD, G. STUPFLER. *Intriguing properties of extreme geometric quantiles*, in "REVSTAT - Statistical Journal", 2015, forthcoming, <https://hal.inria.fr/hal-00865767>
- [26] G. MAZO, S. GIRARD, F. FORBES. *A class of multivariate copulas based on products of bivariate copulas*, in "Journal of Multivariate Analysis", September 2015, vol. 140, pp. 363-376, <https://hal.archives-ouvertes.fr/hal-00910775>
- [27] G. MAZO, S. GIRARD, F. FORBES. *A flexible and tractable class of one-factor copulas*, in "Statistics and Computing", June 2015, forthcoming, <https://hal.archives-ouvertes.fr/hal-00979147>
- [28] G. MAZO, S. GIRARD, F. FORBES. *Weighted least-squares inference for multivariate copulas based on dependence coefficients*, in "ESAIM: Probability and Statistics", October 2015, vol. 19, pp. 746 - 765 [DOI : 10.1051/PS/2015014], <https://hal.archives-ouvertes.fr/hal-00979151>
- [29] 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, <https://hal.inria.fr/hal-01221115>
- [30] S. N. SYLLA, S. GIRARD, A. K. DIONGUE, A. DIALLO, C. SOKHNA. *A classification method for binary predictors combining similarity measures and mixture models*, in "Dependence Modeling", 2015, vol. 3, pp. 240–255, <https://hal.inria.fr/hal-01158043>

- [31] D. WRAITH, F. FORBES. *Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering*, in "Computational Statistics & Data Analysis", 2015, vol. 90, pp. 61-73, <https://hal.archives-ouvertes.fr/hal-01254178>

### Invited Conferences

- [32] J. EL METHNI, L. GARDES, S. GIRARD. *Estimation non-paramétrique de mesures de risque pour des lois conditionnelles à queues lourdes avec application à des extrêmes pluviométriques*, in "Congrès SMAI", Les Karellis, France, June 2015, <https://hal.archives-ouvertes.fr/hal-01168790>
- [33] G. STUPFLER, S. GIRARD. *On the asymptotic behaviour of extreme geometric quantiles*, in "9th International Conference on Extreme Value Analysis", Ann Arbor, United States, 9th International Conference on Extreme Value Analysis, June 2015, <https://hal.archives-ouvertes.fr/hal-01168521>

### International Conferences with Proceedings

- [34] L. AMSALEG, C. OUSSAMA, T. FURON, S. GIRARD, M. E. HOULE, K.-I. KAWARABAYASHI. *Estimating Local Intrinsic Dimensionality*, in "21st Conf. on Knowledge Discovery and Data Mining, KDD2015", Sidney, Australia, ACM, August 2015, <https://hal.inria.fr/hal-01159217>
- [35] A. ARNAUD, F. FORBES, B. LEMASSON, E. L. BARBIER. *Tumor classification and prediction using robust multivariate clustering of multiparametric MRI*, in "International Society for Magnetic Resonance in Medicine", Toronto, Canada, May 2015, <https://hal.archives-ouvertes.fr/hal-01253584>
- [36] F. FORBES, A. FRAU-PASCUAL, P. CIUCIU. *Méthode d'approximation variationnelle pour l'analyse de données d'IRM fonctionnelle acquise par Arterial Spin Labelling*, in "GRETSI", Lyon, France, 2015, <https://hal.archives-ouvertes.fr/hal-01254176>
- [37] A. FRAU-PASCUAL, F. FORBES, P. CIUCIU. *Comparison of Stochastic and Variational Solutions to ASL fMRI Data Analysis*, in "Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015", Munich, Germany, N. NAVAB, J. HORNEGGER, W. M. WELLS III, A. F. FRANGI (editors), Lecture Notes in Computer Science, Springer, October 2015, vol. 9349, n<sup>o</sup> ISBN 978-3-319-24552-2 [DOI : 10.1007/978-3-319-24553-9\_11], <https://hal.archives-ouvertes.fr/hal-01249018>
- [38] A. FRAU-PASCUAL, F. FORBES, P. CIUCIU. *Physiological models comparison for the analysis of ASL FMRI data*, in "12th IEEE International Symposium on Biomedical Imaging, ISBI 2015", New York, United States, 12th IEEE International Symposium on Biomedical Imaging, ISBI 2015, Brooklyn, NY, USA, April 16-19, 2015. IEEE 2015, April 2015, pp. 1348-1351 [DOI : 10.1109/ISBI.2015.7164125], <https://hal.archives-ouvertes.fr/hal-01249014>
- [39] A. FRAU-PASCUAL, F. FORBES, P. CIUCIU. *Variational Physiologically Informed Solution to Hemodynamic and Perfusion Response Estimation from ASL fMRI Data*, in "2015 International Workshop on Pattern Recognition in NeuroImaging", Stanford, CA, United States, 2015 International Workshop on Pattern Recognition in NeuroImaging, Stanford, CA, USA, June 10-12, 2015. IEEE 2015, June 2015, pp. 57-60 [DOI : 10.1109/PRNI.2015.12], <https://hal.archives-ouvertes.fr/hal-01249015>
- [40] 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>

### National Conferences with Proceedings

- [41] A. ARNAUD, F. FORBES, N. COQUERY, E. BARBIER, B. LEMASSON. *Mélanges de lois de Student multivariées généralisées : application à la caractérisation de tumeurs par IRM multiparamétrique*, in "2ème congrès de la SFRMBM (Société Française de Résonance Magnétique en Biologie et Médecine)", Grenoble, France, March 2015, <https://hal.inria.fr/hal-01107483>
- [42] A. ARNAUD, F. FORBES, B. LEMASSON, E. BARBIER. *Paquet R pour l'estimation d'un mélange de lois de Student multivariées à échelles multiples*, in "Quatrièmes Rencontres R", Grenoble, France, June 2015, <https://hal.archives-ouvertes.fr/hal-01253593>
- [43] A. ARNAUD, F. FORBES, B. LEMASSON, E. BARBIER, N. COQUERY. *Mélanges de lois de Student à Échelles Multiples pour la caractérisation de tumeurs par IRM multiparamétrique*, in "47èmes Journées de Statistique de la SFdS", Lille, France, June 2015, <https://hal.archives-ouvertes.fr/hal-01253588>
- [44] S. DOYLE, B. LEMASSON, F. VASSEUR, P. BOURDILLION, F. DUCRAY, J. HONNORAT, L. GUILLON, J. GUYOTAT, C. REMY, F. FORBES, F. COTTON, E. BARBIER, M. DOJAT. *Segmentation des tumeurs cérébrales de bas grade par une approche bayésienne : délimitation manuelle versus automatique*, in "2ème congrès de la SFRMBM (Société Française de Résonance Magnétique en Biologie et Médecine)", Grenoble, France, March 2015, <https://hal.inria.fr/hal-01107520>
- [45] P. FERNIQUE, J.-B. DURAND, Y. GUÉDON. *Détection de motifs disruptifs au sein de plantes : une approche de quotientement/classification d'arborescences*, in "47èmes Journées de Statistique", Lille, France, 47èmes Journées de Statistique, Lille, Société Française de Statistique, June 2015, <https://hal.inria.fr/hal-01240305>
- [46] S. N. SYLLA, S. GIRARD, A. K. DIONGUE, A. DIALLO, C. SOKHNA. *Classification de données binaires via l'introduction de mesures de similarités dans les modèles de mélange*, in "47èmes Journées de Statistique organisées par la Société Française de Statistique", Lille, France, 2015, <https://hal.archives-ouvertes.fr/hal-01168530>

### Conferences without Proceedings

- [47] A. ARNAUD, F. FORBES, B. LEMASSON, E. BARBIER. *Multivariate Multi-scaled Student Distributions : brain tumor characterization from multiparametric MRI*, in "Statistique Mathématique et Applications 2015", Fréjus, France, August 2015, <https://hal.archives-ouvertes.fr/hal-01253595>
- [48] A. CHIANCONE, S. GIRARD, J. CHANUSSOT. *Collaborative Sliced Inverse Regression*, in "20th Young Statisticians Meeting", Vorau, Austria, 2015, <https://hal.archives-ouvertes.fr/hal-01221010>
- [49] S. GIRARD, A. RIVERA, M. STEHLIK, S. TORRES-LEIVA. *Extreme value modelling of some glacial processes in Chilean Andes*, in "International Conference on Risk Analysis", Barcelona, Spain, May 2015, <https://hal.archives-ouvertes.fr/hal-01168766>

### Scientific Books (or Scientific Book chapters)

- [50] P. CIUCIU, F. FORBES, T. VINCENT, L. CHAARI. *Chapter 7 Joint Detection-Estimation in Functional MRI*, in "in Regularization and Bayesian Methods for Inverse Problems in Signal and Image Processing", J.-F. GIOVANNELLI, J. IDIER (editors), John Wiley and Sons, 2015, pp. 160-200, <https://hal.archives-ouvertes.fr/hal-01254177>

- [51] E. H. DEME, S. GIRARD, A. GUILLOU. *Reduced-bias estimator of the Conditional Tail Expectation of heavy-tailed distributions*, in "Mathematical Statistics and Limit Theorems", M. HALLIN (editor), Springer, 2015, pp. 105–123, <https://hal.inria.fr/hal-00823260>
- [52] F. DURANTE, S. GIRARD, G. MAZO. *Copulas based on Marshall–Olkin machinery*, in "Marshall-Olkin Distributions. Advances in Theory and Applications", U. CHERUBINI (editor), Springer Proceedings in Mathematics and Statistics, Springer, 2015, vol. 141, pp. 15–31, <https://hal.archives-ouvertes.fr/hal-01153150>
- [53] S. GIRARD, S. LOUHICHI. *On the strong consistency of the kernel estimator of extreme conditional quantiles*, in "Functional Statistics and Applications", E. OULD SAID (editor), Contributions to Statistics, Springer, 2015, pp. 59–77, <https://hal.inria.fr/hal-01058390>

### Other Publications

- [54] A. ARNAUD, F. FORBES, N. COQUERY, B. LEMASSON, E. BARBIER. *Mélanges de lois de Student multivariées généralisées : application à la caractérisation de tumeurs par IRM multiparamétrique*, April 2015, Forum de la Recherche en Cancérologie Rhône-Alpes Auvergne, Poster, <https://hal.archives-ouvertes.fr/hal-01253576>
- [55] R. AZAÏS, J.-B. DURAND, C. GODIN. *Lossy compression of unordered rooted trees*, December 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01236088>
- [56] L. CHAARI, S. BADILLO, T. VINCENT, G. DEHAENE-LAMBERTZ, F. FORBES, P. CIUCIU. *Hemodynamic-Informed Parcellation of fMRI Data in a Joint Detection Estimation Framework*, November 2015, Submitted to IEEE Transactions on Medical Imaging [DOI : 10.1007/978-3-642-33454-2\_23], <https://hal.archives-ouvertes.fr/hal-01228007>
- [57] 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>
- [58] A. CHIANCONE. *What's wrong with classes? The theory of Knowledge*, February 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01113112>
- [59] A. DAOUIA, S. GIRARD, G. STUPFLER. *Assessing coherent Value-at-Risk and expected shortfall with extreme expectiles*, April 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01142130>
- [60] S. GIRARD, J. SARACCO. *Supervised and unsupervised classification using mixture models*, December 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01241818>

### References in notes

- [61] 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>
- [62] P. EMBRECHTS, C. KLÜPPELBERG, T. MIKOSH. *Modelling Extremal Events*, Applications of Mathematics, Springer-Verlag, 1997, vol. 33

- [63] F. FERRATY, P. VIEU. *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, 2006
- [64] 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 Cergy-Pontoise, octobre 1996
- [65] 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>
- [66] K. LI. *Sliced inverse regression for dimension reduction*, in "Journal of the American Statistical Association", 1991, vol. 86, pp. 316–327
- [67] R. NELSEN. *An introduction to copulas*, Lecture Notes in Statistics, Springer-Verlag, New-York, 1999, vol. 139
- [68] J. SIMOLA, J. SALOJÄRVI, I. KOJO. *Using hidden Markov model to uncover processing states from eye movements in information search tasks*, in "Cognitive Systems Research", Oct 2008, vol. 9, n<sup>o</sup> 4, pp. 237-251