



IN PARTNERSHIP WITH:
CNRS

**Institut polytechnique de
Grenoble**

**Université Joseph Fourier
(Grenoble)**

Activity Report 2013

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

1. Members	1
2. Overall Objectives	1
2.1. Introduction	1
2.2. Highlights of the Year	2
2.2.1. European project HUMAVIPS.	2
2.2.2. Best Paper Award at IEEE MMSP'13.	2
3. Research Program	2
3.1. Mixture models	2
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	5
3.3.3. Dimension reduction	6
4. Application Domains	6
4.1. Image Analysis	6
4.2. Biology, Environment and Medicine	6
5. Software and Platforms	6
5.1. The LOCUS software	6
5.2. The P-LOCUS software	7
5.3. The PyHRF software	7
6. New Results	8
6.1. Mixture models	8
6.1.1. Parameter estimation in the heterogeneity linear mixed model	8
6.1.2. Taking into account the curse of dimensionality	9
6.1.3. Mixture modelling using skewed multivariate heavy tailed distributions with variable amounts of tailweight	9
6.1.4. High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables	9
6.1.5. Acoustic space learning via variational EM for Sound-Source Separation and Localization	10
6.2. Statistical models for Neuroscience	10
6.2.1. Hemodynamically informed parcellation of cerebral fMRI data	10
6.2.2. Variational variable selection to assess experimental condition relevance in event-related fMRI	11
6.2.3. Bayesian Joint Detection-Estimation of cerebral vasoreactivity from ASL fMRI data	11
6.2.4. Physiologically-inspired Bayesian analysis of BOLD and ASL fMRI data	11
6.3. Markov models	12
6.3.1. Spatial modelling of plant diversity from high-throughput environmental DNA sequence data	12
6.3.2. Modelling multivariate counts with graphical Markov models.	12
6.3.3. Statistical characterization of tree structures based on Markov tree models and multitype branching processes, with applications to tree growth modelling.	13
6.3.4. Statistical characterization of the alternation of flowering in fruit tree species	14
6.4. Semi and non-parametric methods	14
6.4.1. Modelling extremal events	14
6.4.2. Conditional extremal events	15
6.4.3. Estimation of extreme risk measures	15
6.4.4. Multivariate extremal events	15
6.4.5. Level sets estimation	16

6.4.6.	Retrieval of Mars surface physical properties from OMEGA hyperspectral images.	16
6.4.7.	High-dimensional change-point detection with sparse alternatives	17
6.4.8.	Yield Improvement by the Redundancy Method for Component Calibration	17
7.	Partnerships and Cooperations	18
7.1.	Regional Initiatives	18
7.2.	National Initiatives	18
7.2.1.	Competitvity Clusters	18
7.2.2.	Inria project HEROES	18
7.3.	European Initiatives	18
7.4.	International Initiatives	19
7.5.	International Research Visitors	19
8.	Dissemination	20
8.1.	Scientific Animation	20
8.2.	Teaching - Supervision - Juries	20
8.2.1.	Teaching	20
8.2.2.	Supervision	21
8.2.3.	Juries	21
9.	Bibliography	22

Project-Team MISTIS

Keywords: Stochastic Models, Machine Learning, Data Analysis, Image Processing, Statistical Methods

Creation of the Project-Team: 2008 January 01.

1. Members

Research Scientists

Florence Forbes [Team leader, Inria, Senior Researcher, HdR]

Stéphane Girard [Inria, Researcher, HdR]

Faculty Members

Jean-Baptiste Durand [Grenoble INP, Associate Professor]

Marie-José Martinez [Univ. Grenoble II, Associate Professor]

Engineers

Flor Vasseur [Inria]

Senan Doyle [Inria, until Nov 2013]

PhD Students

Christine Bakhous [Inria, Cordi-S]

Alessandro Chiancone [Univ. Grenoble I, from Oct 2013]

Jonathan El-Methni [Univ. Grenoble I, until Sep 2013]

Aina Frau Pascual [Inria, Cordi-S, from Oct 2013]

Gildas Mazo [Inria, Cordi-S]

Post-Doctoral Fellows

Thomas Vincent [Inria]

Darren Wraith [Inria, until Apr 2013]

Angelika Studeny [Inria]

Farida Enikeeva [Ensimag]

Visiting Scientists

El-Hadji Deme [Univ. St Louis, Senegal, Post-doc, from Sept 2013]

Alexander Nazin [Aca. Sciences, Moscou, from May 2013 until Oct 2013]

Seydou-Nourou Sylla [IRD, PhD]

Administrative Assistant

Françoise de Coninck [Inria]

Others

Van Trung Pham [Inria, Intern, from Apr 2013 until Sep 2013]

Jennifer Sloboda [Inria, Intern, from May 2013 until Aug 2013]

2. Overall Objectives

2.1. Introduction

The MISTIS team aims at developing statistical methods for dealing with complex problems or data. Our applications relate mainly to image processing and spatial data problems with some applications in environment, biology and medicine. Our approach is based on the statement that complexity can be handled by working up from simple local assumptions in a coherent way, defining a structured model, and that is the key to modelling, computation, inference and interpretation. The methods we focus on involve mixture models, Markov models, and, more generally, hidden structure models identified by deterministic or 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 statistical methodology (finite mixture analysis, hidden Markov models, random effect models, etc). 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. The focus is on functions describing curves or surfaces or more generally manifolds rather than real valued parameters. This can be interesting in image processing for instance where it can be difficult to introduce parametric models that are general enough (e.g. for contours).

2.2. Highlights of the Year

2.2.1. European project HUMAVIPS.

The European project **HUMAVIPS** – Humanoids with Auditory and Visual Abilities in Populated Spaces – is a 36-month FP7 STREP project coordinated by Radu Horaud and which started in 2010. The project addressed multimodal perception and cognitive issues associated with the computational development of a social robot. The objective was to endow humanoid robots with audiovisual (AV) abilities: exploration, recognition, and interaction, such that they exhibit adequate behavior when dealing with a group of people. Research and technological developments emphasized the role played by multimodal perception within principled models of human-robot interaction and of humanoid behavior. The HUMAVIPS project was successfully terminated in January 2013.

An article about *Integrating Smart Robots into Society* refers to HUMAVIPS. The article stresses the role of cognition in human-robot interaction and refers to HUMAVIPS as one of the FP7 projects that has paved the way towards the concept of audio-visual robotics. The article was published in **HORIZON**, which is Europe's Research & Innovation Magazine.

2.2.2. Best Paper Award at IEEE MMSP'13.

The paper addresses the problem of aligning visual and auditory data using a sensor that is composed of a camera-pair and a microphone-pair. The original contribution of the paper is a method for audio-visual data aligning through estimation of the 3D positions of the microphones in the visual centred coordinate frame defined by the stereo camera-pair. Please consult http://www.mmsp2013.org/mmsp2013_awards.php and .

BEST PAPER AWARD :

[41] **Alignment of Binocular-Binaural Data Using a Moving Audio-Visual Target in MMSP 2013 - IEEE International Workshop on Multimedia Signal Processing.** V. KHALIDOV, F. FORBES, R. HORAUD.

3. Research Program

3.1. Mixture models

Participants: Angelika Studeny, Thomas Vincent, Christine Bakhous, Senan James Doyle, Jean-Baptiste Durand, Florence Forbes, Aina Frau Pascual, Alessandro Chiancone, Stéphane Girard, Marie-José Martinez, Darren Wraith.

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

In a first approach, we consider statistical parametric models, θ being the parameter, possibly multi-dimensional, usually unknown and to be estimated. We consider cases where the data naturally divides into observed data $y = y_1, \dots, y_n$ 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: Angelika Studeny, Thomas Vincent, Christine Bakhous, Senan James Doyle, Jean-Baptiste Durand, Florence Forbes, Darren Wraith.

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: El-Hadji Deme, Jonathan El-Methni, Stéphane Girard, Gildas Mazo, Farida Enikeeva, Seydou-Nourou Sylla.

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 [66]. 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 [71] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [65], 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).

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 ℓ 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 ρ is, the slower the convergence in (3) and the more difficult the estimation of extreme quantiles.

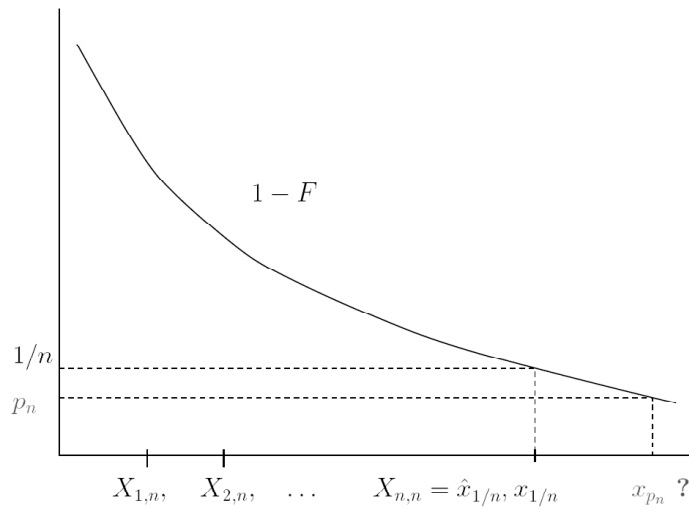


Figure 1. The curve represents the survival function $x \rightarrow P(X > x)$. The $1/n$ -quantile is estimated by the maximum observation so that $\hat{x}_{1/n} = X_{n,n}$. As illustrated in the figure, to estimate p_n -quantiles with $p_n < 1/n$, it is necessary to extrapolate beyond the maximum observation.

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 [67]. 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 [64]. 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 [71].

4. Application Domains

4.1. Image Analysis

Participants: Christine Bakhous, Senan James Doyle, Aina Frau Pascual, Thomas Vincent, Florence Forbes, Stéphane Girard, Darren Wraith, Flor Vasseur, Alessandro Chiancone, Farida Enikeeva.

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. In the context of the ANR MDCO project Vahine, we worked on hyperspectral multi-angle 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: Thomas Vincent, Christine Bakhous, Aina Frau-Pascual, Senan James Doyle, Florence Forbes, Stéphane Girard, Jonathan El-Methni, Gildas Mazo, Angelika Studeny, Seydou-Nourou Sylla, Marie-José Martinez, Jean-Baptise 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 population genetics and 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. Software and Platforms

5.1. The LOCUS software

Participants: Florence Forbes, Senan James Doyle.

Joint work with: Michel Dojat from Grenoble Institute of Neuroscience and Benoit Scherrer from Harvard Medical School, Boston, MA, USA.

From brain MR images, neuroradiologists are able to delineate tissues such as grey matter and structures such as Thalamus and damaged regions. This delineation is a common task for an expert but unsupervised segmentation is difficult due to a number of artefacts. The LOCUS software (<http://locus.gforge.inria.fr>) automatically perform this segmentation for healthy brains. An image is divided into cubes on each of which a statistical model is applied. This provides a number of local treatments that are then integrated to ensure consistency at a global level, resulting in low sensitivity to artifacts. The statistical model is based on a Markovian approach that enables to capture the relations between tissues and structures, to integrate a priori anatomical knowledge and to handle local estimations and spatial correlations.

The LOCUS software has been developed in the context of a collaboration between Mistis, a computer science team (Magma, LIG) and a Neuroscience methodological team (the Neuroimaging team from Grenoble Institut of Neurosciences, INSERM). This collaboration resulted over the period 2006-2008 into the PhD thesis of B. Scherrer (advised by C. Garbay and M. Dojat) and in a number of publications. In particular, B. Scherrer received a "Young Investigator Award" at the 2008 MICCAI conference.

The originality of this work comes from the successful combination of the teams respective strengths i.e. expertise in distributed computing, in neuroimaging data processing and in statistical methods.

5.2. The P-LOCUS software

Participants: Florence Forbes, Senan James Doyle, Flor Vasseur.

Joint work with: Michel Dojat.

The Locus software was extended to address the delineation of lesions in pathological brains. Its extension P-LOCUS (<http://p-locus.com>) for lesion detection was realized by S. Doyle with financial support from Gravit with the goal to create a Start-up. 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. Its originality comes from:

- it is fully automatic: no external user interaction and no training data required
- the possibility to combine information from several images (MR sequences)
- a statistical Bayesian framework for robustness to image artefacts and a priori knowledge incorporation
- a voxel-based clustering technique that uses Markov random fields (MRF) incorporating information about neighboring voxels for spatial consistency and robustness to imperfect image features (noise).
- the possibility to select and incorporate relevant a priori knowledge via different atlases, e.g. tissue and vascular territory atlases
- a fully integrated preprocessing steps and lesion ROI identification

P-LOCUS software was presented at various conferences and used for the BRATS Challenge on tumor segmentation organized as a satellite challenge of the Miccai conference in Nagoya, Japan. A paper submitted to IEEE trans. on Medical Imaging reports the challenge results [62].

5.3. The PyHRF software

Participants: Christine Bakhous, Florence Forbes, Thomas Vincent.

Joint work with: Philippe Ciuciu and Solveig Badillo from Parietal Team Inria and CEA NeuroSpin, Lotfi Chaari and Laurent Risser from Toulouse University.

As part of fMRI data analysis, the PyHRF package (<http://pyhrf.org>) 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. With respect to the sole detection issue (i), the classical voxelwise GLM procedure is also available through NIPY, whereas Finite Impulse Response (FIR) and temporally regularized FIR models are implemented to deal with the HRF estimation concern (ii). Several parcellation tools are also integrated such as spatial and functional clusterings. Parcellations may be used for spatial averaging prior to FIR/RFIR analysis or to specify the spatial support of the HRF estimates in the JDE approach. These analysis procedures can be applied either to volumic data sets or to data projected onto the cortical surface. For validation purpose, this package is shipped with artificial and real fMRI data sets. To cope with the high computational needs for inference, PyHRF handles distributing computing by exploiting cluster units as well as multiple cores computers. Finally, a dedicated viewer is available which handles n -dimensional images and provides suitable features for exploring whole brain hemodynamics (display of time series, maps, ROI mask overlay). A paper under revision for *Frontiers in Neuroinformatics* gives more details on the current PyHRF functionalities.

6. New Results

6.1. Mixture models

6.1.1. *Parameter estimation in the heterogeneity linear mixed model*

Participant: Marie-José Martinez.

Joint work with: Emma Holian (National University of Ireland, Galway)

In studies where subjects contribute more than one observation, such as in longitudinal studies, linear mixed models have become one of the most used techniques to take into account the correlation between these observations. By introducing random effects, mixed models allow the within-subject correlation and the variability of the response among the different subjects to be taken into account. However, such models are based on a normality assumption for the random effects and reflect the prior belief of homogeneity among all the subjects. To relax this strong assumption, Verbeke and Lesaffre (1996) proposed the extension of the classical linear mixed model by allowing the random effects to be sampled from a finite mixture of normal distributions with common covariance matrix. This extension naturally arises from the prior belief of the presence of unobserved heterogeneity in the random effects population. The model is therefore called the heterogeneity linear mixed model. Note that this model does not only extend the assumption about the random effects distribution, indeed, each component of the mixture can be considered as a cluster containing a proportion of the total population. Thus, this model is also suitable for classification purposes.

Concerning parameter estimation in the heterogeneity model, the use of the EM-algorithm, which takes into account the incomplete structure of the data, has been considered in the literature. Unfortunately, the M-step in the estimation process is not available in analytic form and a numerical maximisation procedure such as Newton-Raphson is needed. Because deriving such a procedure is a non-trivial task, Komarek et al. (2002) proposed an approximate optimization. But this procedure proved to be very slow and limited to small samples due to requiring manipulation of very large matrices and prohibitive computation.

To overcome this problem, we have proposed in [28], [52] an alternative approach which consists of fitting directly an equivalent mixture of linear mixed models. Contrary to the heterogeneity model, the M-step of the EM-algorithm is tractable analytically in this case. Then, from the obtained parameter estimates, we can easily obtain the parameter estimates in the heterogeneity model.

6.1.2. *Taking into account the curse of dimensionality*

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

Joint work with: C. Bouveyron (Univ. Paris 1), M. Fauvel (ENSAT Toulouse) and J. Chaussoot (Gipsa-lab and Grenoble-INP)

In the PhD work of Charles Bouveyron (co-advised by Cordelia Schmid from the Inria LEAR team) [64], we propose 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 new 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.

6.1.3. *Mixture modelling using skewed multivariate heavy tailed distributions with variable amounts of tailweight*

Participants: Florence Forbes, Darren Wraith.

Clustering concerns the assignment of each of N , possibly multidimensional, observations y_1, \dots, y_N to one of K groups. A popular way to approach this task is via a parametric finite mixture model. While the vast majority of the work on such mixtures has been based on Gaussian mixture models in many applications the tails of normal distributions are shorter than appropriate or parameter estimations are affected by atypical observations (outliers). The family of location and scale mixtures of Gaussians has the ability to generate a number of flexible distributional forms. It nests as particular cases several important asymmetric distributions like the Generalised Hyperbolic distribution. The Generalised Hyperbolic distribution in turn nests many other well known distributions such as the Normal Inverse Gaussian (NIG) whose practical relevance has been widely documented in the literature. In a multivariate setting, we propose to extend the standard location and scale mixture concept into a so called multiple scaled framework which has the advantage of allowing different tail and skewness behaviours in each dimension of the variable space with arbitrary correlation between dimensions. The approach builds upon, and develops further, previous work on scale mixtures of Gaussians [25]. Estimation of the parameters is provided via an EM algorithm with a particular focus on NIG distributions. Inference is then extended to cover the case of mixtures of such multiple scaled distributions for application to clustering. Assessments on simulated and real data confirm the gain in degrees of freedom and flexibility in modelling data of varying tail behaviour and directional shape.

6.1.4. *High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables*

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

In this work we address the problem of approximating high-dimensional data with a low-dimensional representation. We make the following contributions. We propose an inverse regression method which exchanges the roles of input and response, such that the low-dimensional variable becomes the regressor, and which is tractable. We introduce a mixture of locally-linear probabilistic mapping model that starts with estimating the parameters of inverse regression, and follows with inferring closed-form solutions for the forward parameters of the high-dimensional regression problem of interest. Moreover, we introduce a partially-latent paradigm, such that the vector-valued response variable is composed of both observed and latent entries, thus being able to deal with data contaminated by experimental artifacts that cannot be explained with noise models. The proposed probabilistic formulation could be viewed as a latent-variable augmentation of regression. We devise expectation-maximization (EM) procedures based on a data augmentation strategy which facilitates the maximum-likelihood search over the model parameters. We propose two augmentation schemes and we describe in detail the associated EM inference procedures that may well be viewed as generalizations of a number of EM regression, dimension reduction, and factor analysis algorithms. The proposed framework is validated with both synthetic and real data. We provide experimental evidence that our method outperforms several existing regression techniques.

6.1.5. *Acoustic space learning via variational EM for Sound-Source Separation and Localization*

Participant: Florence Forbes.

Joint work with: Antoine Deleforge and Radu Horaud from the Inria Perception team.

In this paper we address the problems of modeling the acoustic space generated by a full-spectrum sound source and of using the learned model for the localization and separation of multiple sources that simultaneously emit sparse-spectrum sounds. We lay theoretical and methodological grounds in order to introduce the *binaural manifold* paradigm. We perform an in-depth study of the latent low-dimensional structure of the high-dimensional interaural spectral data, based on a corpus recorded with a human-like audiomotor robot head. A non-linear dimensionality reduction technique is used to show that these data lie on a two-dimensional (2D) smooth manifold parameterized by the motor states of the listener, or equivalently, the sound source directions. We propose a *probabilistic piecewise affine mapping* model (PPAM) specifically designed to deal with high-dimensional data exhibiting an intrinsic piecewise linear structure. We derive a closed-form expectation-maximization (EM) procedure for estimating the model parameters, followed by Bayes inversion for obtaining the full posterior density function of a sound source direction. We extend this solution to deal with missing data and redundancy in real world spectrograms, and hence for 2D localization of natural sound sources such as speech. We further generalize the model to the challenging case of multiple sound sources and we propose a variational EM framework. The associated algorithm, referred to as *variational EM for source separation and localization* (VESSL) yields a Bayesian estimation of the 2D locations and time-frequency masks of all the sources. Comparisons of the proposed approach with several existing methods reveal that the combination of acoustic-space learning with Bayesian inference enables our method to outperform state-of-the-art methods.

6.2. Statistical models for Neuroscience

6.2.1. *Hemodynamically informed parcellation of cerebral fMRI data*

Participants: Florence Forbes, Aina Frau-Pascual, Thomas Vincent.

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

Standard detection of evoked brain activity in functional MRI (fMRI) relies on a fixed and known shape of the impulse response of the neurovascular coupling, namely the hemodynamic response function (HRF). To cope with this issue, the joint detection-estimation (JDE) framework has been proposed. This formalism enables to estimate a HRF per region but for doing so, it assumes a prior brain partition (or parcellation) regarding hemodynamic territories (eg. [14]). This partition has to be accurate enough to recover accurate HRF shapes but has also to overcome the detection-estimation issue: the lack of hemodynamics information in the non-active positions. During the internship of A. Frau Pascual at Neurospin, we proposed an hemodynamically-based parcellation, consisting first of a feature extraction step, followed by a Gaussian Mixture-based

parcellation, which considers the injection of the activation levels in the parcellation process, in order to overcome the detection-estimation issue and find the underlying hemodynamics. The work has been submitted to the ICASSP conference in 2014.

6.2.2. *Variational variable selection to assess experimental condition relevance in event-related fMRI*

Participants: Florence Forbes, Christine Bakhous, Lotfi Chaari, Thomas Vincent, Farida Enikeeva.

Joint work with: Michel Dojat (Grenoble Institute of Neuroscience) and Philippe Ciuciu from Neurospin, CEA in Saclay.

Brain functional exploration investigates the nature of neural processing following cognitive or sensory stimulation. This goal is not fully accounted for in most functional Magnetic Resonance Imaging (fMRI) analysis which usually assumes that all delivered stimuli possibly generate a BOLD response everywhere in the brain although activation is likely to be induced by only some of them in specific brain regions. Generally, criteria are not available to select the relevant conditions or stimulus types (e.g. visual, auditory, etc.) prior to activation detection and the inclusion of irrelevant events may degrade the results, particularly when the Hemodynamic Response Function (HRF) is jointly estimated as in the JDE framework mentioned in the previous section. To face this issue, we propose an efficient variational procedure that automatically selects the conditions according to the brain activity they elicit. It follows an improved activation detection and local HRF estimation that we illustrate on synthetic and real fMRI data. This approach is an alternative to our previous approach based on Monte-Carlo Markov Chain (MCMC) inference [63]. Corresponding papers [31], [45]. A synthesis can also be found in the PhD manuscript of C. Bakhous (Grenoble University, December 2013) [11].

6.2.3. *Bayesian Joint Detection-Estimation of cerebral vasoreactivity from ASL fMRI data*

Participants: Florence Forbes, Thomas Vincent.

In the context of ARC AINSI project, joint work with: Philippe Ciuciu from Neurospin, CEA in Saclay.

Functional MRI (fMRI) is the method of choice to non-invasively probe cerebral activity evoked by a set of controlled experimental conditions. A rising fMRI modality is Arterial Spin Labeling (ASL) which enables to quantify the cerebral perfusion, namely the cerebral blood flow (CBF) and emerges as a more direct biomarker of neuronal activity than the standard BOLD (Blood Oxygen Level Dependent) fMRI.

Although the study of cerebral vasoreactivity using fMRI is mainly conducted through the BOLD fMRI modality (see the two previous sections), owing to its relatively high signal-to-noise ratio (SNR), ASL fMRI provides a more interpretable measure of cerebral vasoreactivity than BOLD fMRI. Still, ASL suffers from a low SNR and is hampered by a large amount of physiological noise. Our contribution, described in [43], [44] aims at improving the recovery of the vasoreactive component from the ASL signal. To this end, a Bayesian hierarchical model is proposed, enabling the recovery of perfusion levels as well as fitting their dynamics. On a single-subject ASL real data set involving perfusion changes induced by hypercapnia, the approach is compared with a classical GLM-based analysis. A better goodness-of-fit is achieved, especially in the transitions between baseline and hypercapnia periods. Also, perfusion levels are recovered with higher sensitivity and show a better contrast between gray- and white matter.

6.2.4. *Physiologically-inspired Bayesian analysis of BOLD and ASL fMRI data*

Participants: Florence Forbes, Thomas Vincent, Jennifer Sloboda.

In the context of ARC AINSI project, joint work with: Philippe Ciuciu from Neurospin, CEA in Saclay.

The ASL modality is most commonly used as a static measure where the average perfusion is computed over a volume sequence lasting several minutes. Recently, ASL has been used in functional activation protocols and hence gives access to a dynamic measure of perfusion, namely the variations of CBF which are elicited by specific tasks. ASL MRI mainly consists of acquiring pairs of control and label images and looking at the average control-label difference. The Signal-to-Noise Ratio (SNR) of this difference is very low so that several hundreds of image pairs need to be acquired, thus increasing significantly the time spent by the

subject in the scanner and making the acquisition very sensitive to the patient's movement. In addition, this averaging requires that the perfusion signal is at a steady state, limiting the scope of fMRI task experiments to baseline perfusion measurements or long block designs. In contrast, it is highly desirable to measure change in perfusion due to an effect of interest in activation paradigms from event-related designs. It is technically possible to collect event-related ASL data but most approaches to functional ASL data analysis use a standard linear model (GLM-based) formulation with regressors encoding differences in control/tag scans and both ASL and BOLD activation signals being associated with the same canonical response function. The canonical hemodynamic response function (HRF) is generally used although it has been calibrated on BOLD experiments only, thus reflecting simultaneous variations of CBF, cerebral blood volume (CBV) and cerebral oxygen consumption (CMRO₂). In contrast, the perfusion signal only reflects variation in CBF so that the associated response, the perfusion response function (PRF), is likely to differ from the HRF. In the internship proposal of Jennifer Sloboda, we proposed to recover both a hemodynamic (BRF for BOLD response function) and a perfusion (PRF) response functions from event-related functional ASL data. To do so, a joint detection estimation (JDE) formalism was used. In the BOLD context, the JDE framework has proven to successfully extract the HRF while also performing activation detection. We had recently extended this formalism (see Section 6.2.3 and [43], [44]) to model an additional perfusion component linked to the BOLD one through a common activation detection. The main issue addressed then was to characterize the link between BOLD and perfusion components. To establish this link, we proposed a methodological axis which consists of developing a physiologically-inspired approach. To do so, dynamical non-linear equations available in physiological models were linearized and approximated in a parsimonious way so as to establish prior relations between the perfusion and BOLD responses which can be injected in our Bayesian setting. The inference of the initial model is currently done through a Markov Chain Monte Carlo approach but a Variational Expectation-Maximization implementation is also conceivable. As such, the tasks were two-fold: (1) investigate the physiological model and (2) inject it into the JDE setting. Investigation of the physiological model allows for: (1) creation of artificial fMRI data, (2) investigation of the relationship between physiological changes and the resulting simulated BOLD or ASL signal, and (3) characterization of the link between BOLD and perfusion responses. Injection of the physiologically inspired prior into the JDE model, is to (1) improve perfusion response recovery, (2) determine physiologically quantified units to the JDE recovered values. This work is going to serve as a preliminary investigation into the incorporation of physiological information in the Bayesian JDE setting from which to determine the trajectory of future model developments.

6.3. Markov models

6.3.1. *Spatial modelling of plant diversity from high-throughput environmental DNA sequence data*

Participants: Florence Forbes, Angelika Studeny.

This is joint work with: Eric Coissac and Pierre Taberlet from LECA (Laboratoire d'Ecologie Alpine) and Alain Viari from Inria team Bamboo.

This work [48] considers a statistical modelling approach to investigate spatial cross-correlations between species in an ecosystem. A special feature is the origin of the data from high-throughput environmental DNA sequencing of soil samples. Here we use data collected at the Nourague CNRS Field Station in French Guiana. We describe bivariate spatial relationships in these data by a separable linear model of coregionalisation and estimate a cross-correlation parameter. Based on this estimate, we visualise plant taxa co-occurrence pattern in form of 'interaction graphs' which can be interpreted in terms of ecological interactions. Limitations of this approach are discussed along with possible alternatives in [48].

6.3.2. *Modelling multivariate counts with graphical Markov models.*

Participants: Jean-Baptiste Durand, Florence Forbes, Marie-José Martinez, Angelika Studeny.

Joint work with: Pierre Fernique (Montpellier 2 University, CIRAD and Inria Virtual Plants), Yann Guédon (CIRAD and Inria Virtual Plants) and Iragâël Joly (INRA-GAEL and Grenoble INP).

Multivariate count data are defined as the number of items of different categories issued from sampling within a population, which individuals are grouped into categories. The analysis of multivariate count data is a recurrent and crucial issue in numerous modelling problems, particularly in the fields of biology and ecology (where the data can represent, for example, children counts associated with multitype branching processes), sociology and econometrics. Denoting by K the number of categories, multivariate count data analysis relies on modelling the joint distribution of the K -dimensional random vector $N = (N_0, \dots, N_{K-1})$ with discrete components. Our work focused on I) Identifying categories that appear simultaneously, or on the contrary that are mutually exclusive. This was achieved by identifying conditional independence relationships between the K variables; II) Building parsimonious parametric models consistent with these relationships; III) Characterizing and testing the effects of covariates on the distribution of N , particularly on the dependencies between its components.

Our context of application was characterised by zero-inflated, often right skewed marginal distributions. Thus, Gaussian and Poisson distributions were not *a priori* appropriate. Moreover, the multivariate histograms typically had many cells, most of which were empty. Consequently, nonparametric estimation was not efficient.

To achieve these goals, we proposed an approach based on graphical probabilistic models (Koller & Friedman, 2009 [70]) to represent the conditional independence relationships in N , and on parametric distributions to ensure model parsimony [51]. The considered graphs were partially directed, so as to represent both marginal independence relationships and cyclic dependencies between quadruplets of variables (at least).

Graph search was achieved by a stepwise approach, issued from unification of previous algorithms presented in Koller & Friedman (2009) for DAGs: Hill climbing, greedy search, first ascent and simulated annealing. The search algorithm was improved by taking into account our parametric distribution assumptions, which led to caching overlapping graphs at each step. An adaptation to PDAGs of graph search algorithms for DAGs was developed, by defining new operators specific to PDAGs.

Comparisons between different algorithms in the literature for directed and undirected graphical models was performed on simulated datasets to: (i) Assess gain in speed induced by caching; (ii) Compare the graphs obtained under parametric and nonparametric distributions assumptions; (iii) Compare different strategies for graph initialization. Strategies based on several random graphs were compared to those based on a fast estimation of an undirected graph, assumed to be the moral graph.

First results were obtained in modelling individual daily activity program [50] and interactions between flowering and vegetative growth in plants (see sections below).

6.3.3. *Statistical characterization of tree structures based on Markov tree models and multitype branching processes, with applications to tree growth modelling.*

Participant: Jean-Baptiste Durand.

Joint work with: Pierre Fernique (Montpellier 2 University and CIRAD) and Yann Guédon (CIRAD), Inria Virtual Plants.

The quantity and quality of yields in fruit trees is closely related to processes of growth and branching, which determine ultimately the regularity of flowering and the position of flowers. Flowering and fruiting patterns are explained by statistical dependence between the nature of a parent shoot (*e.g.* flowering or not) and the quantity and natures of its children shoots – with potential effect of covariates. Thus, better characterization of patterns and dependences is expected to lead to strategies to control the demographic properties of the shoots (through varietal selection or crop management policies), and thus to bring substantial improvements in the quantity and quality of yields.

Since the connections between shoots can be represented by mathematical trees, statistical models based on multitype branching processes and Markov trees appear as a natural tool to model the dependencies of interest. Formally, the properties of a vertex are summed up using the notion of vertex state. In such models, the numbers of children in each state given the parent state are modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious distributions. We developed an approach based on probabilistic graphical models (see Section 6.3.2) to identify and exploit properties of conditional

independence between numbers of children in different states, so as to simplify the specification of their joint distribution [51], [32].

This work was carried out in the context of Pierre Fernique's first year of PhD (Montpellier 2 University and CIRAD). It was applied to model dependencies between short or long, vegetative or flowering shoots in apple trees. The results highlighted contrasted patterns related to the parent shoot state, with interpretation in terms of alternation of flowering (see paragraph 6.3.4). It was also applied to the analysis of the connections between cyclic growth and flowering of mango trees [32]. This work will be continued during Pierre Fernique's PhD thesis, with extensions to other fruit tree species and other parametric discrete multivariate families of distributions, including covariates and mixed effects.

6.3.4. *Statistical characterization of the alternation of flowering in fruit tree species*

Participant: Jean-Baptiste Durand.

Joint work with: Jean Peyhardi and Yann Guédon (Mixed Research Unit DAP, Virtual Plants team), Baptiste Guitton, Yan Holtz and Evelyne Costes (DAP, AFEF team), Catherine Trottier (Montpellier University)

A first study was performed to characterize genetic determinisms of the alternation of flowering in apple tree progenies [37], [21]. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot or AS, which is the portions of stem that were grown during the same year). Two replications of each genotype were available.

Indices were proposed to characterize alternation at tree scale. The difficulty is related to early detection of alternating genotypes, in a context where alternation is often concealed by a substantial increase of the number of flowers over consecutive years. To separate correctly the increase of the number of flowers due to aging of young trees from alternation in flowering, our model relied on a parametric hypothesis for the trend (fixed slopes specific to genotype and random slopes specific to replications), which translated into mixed effect modelling. Then, different indices of alternation were computed on the residuals. Clusters of individuals with contrasted patterns of bearing habits were identified.

To model alternation of flowering at AS scale, a second-order Markov tree model was built. Its transition probabilities were modelled as generalized linear mixed models, to incorporate the effects of genotypes, year and memory of flowering for the Markovian part, with interactions between these components.

Asynchronism of flowering at AS scale was assessed using an entropy-based criterion. The entropy allowed for a characterisation of the roles of local alternation and asynchronism in regularity of flowering at tree scale.

Moreover, our models highlighted significant correlations between indices of alternation at AS and individual scales.

This work was extended by the Master 2 internship of Yan Holtz, supervised by Evelyne Costes and Jean-Baptiste Durand. New progenies were considered, and a methodology based on a lighter measurement protocol was developed and assessed. It consisted in assessing the accuracy of approximating the indices computed from measurements at tree scale by the same indices computed as AS scale. The approximations were shown sufficiently accurate to provide an operational strategy for apple tree selection.

As a perspective of this work, patterns in the production of children ASs (numbers of flowering and vegetative children) depending on the type of the parent AS must be analyzed using branching processes and different types of Markov trees, in the context of Pierre Fernique's PhD Thesis (see paragraph 6.3.3).

6.4. Semi and non-parametric methods

6.4.1. *Modelling extremal events*

Participants: Stéphane Girard, El-Hadji Deme.

Joint work with: L. Gardes (Univ. Strasbourg) and E. Deme (Univ. Gaston Berger, Sénégal)

We are working on the estimation of the second order parameter ρ (see paragraph 3.3.1). We proposed a new family of estimators encompassing the existing ones (see for instance [69], [68]). This work is in collaboration with El-Hadji Deme who obtained a grant (IBNI price) to work within the Mistis team on extreme-value statistics. The results are published in [18].

In addition to this work, we have established a review on the Weibull-tail distributions [29].

6.4.2. Conditional extremal events

Participants: Stéphane Girard, Gildas Mazo, Jonathan El-Methni.

Joint work with: L. Gardes (Univ. Strasbourg) and A. Daouia (Univ. Toulouse I and Univ. Catholique de Louvain)

The goal of the PhD thesis of Alexandre Lekina 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 [66] 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 [16].

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.

6.4.3. Estimation of extreme risk measures

Participants: Stéphane Girard, Jonathan El-Methni, El-Hadji Deme.

Joint work with: L. Gardes and A. Guillou (Univ. Strasbourg)

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: $\text{VaR}(\alpha)$ 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 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. It is frequently encountered in financial investment or in the insurance industry. In [36], we have established the asymptotic properties of the classical 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 [36]. Similar developments have been achieved in the case of the Proportional Hazard Premium measure of risk [19]. In [22], we study the situation where some covariate information is available. We thus have to deal with conditional extremes (see paragraph 6.4.2). We also proposed a new risk measure (called the Conditional Tail Moment) which encompasses various risk measures, such as the CTE, as particular cases.

6.4.4. Multivariate extremal events

Participants: Stéphane Girard, Gildas Mazo, Florence Forbes, Van Trung Pham.

Joint work with: C. Amblard (TimB in TIMC laboratory, Univ. Grenoble I) and L. Menneteau (Univ. Montpellier II)

Copulas are a useful tool to model multivariate distributions [72]. At first, we developed an extension of some particular copulas [1]. It followed a new class of bivariate copulas defined on matrices [56] and some analogies have been shown between matrix and copula properties.

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 [61]. 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. A pairwise moment-based inference procedure has also been proposed and the asymptotic normality of the corresponding estimator has been established [53].

6.4.5. Level sets estimation

Participant: Stéphane Girard.

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

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 [60] and thus should not be used to detect outliers.

In collaboration with A. Daouia, we investigate the application of such methods in econometrics [17]: A new characterization of partial boundaries of a free disposal multivariate support is introduced by making use of large quantiles of a simple transformation of the underlying multivariate distribution. Pointwise empirical and smoothed estimators of the full and partial support curves are built as extreme sample and smoothed quantiles. The extreme-value theory holds then automatically for the empirical frontiers and we show that some fundamental properties of extreme order statistics carry over to Nadaraya's estimates of upper quantile-based frontiers.

In collaboration with G. Stupfler and A. Guillou, new estimators of the boundary are introduced. The regression is performed on the whole set of points, the selection of the "highest" points being automatically performed by the introduction of high order moments [26], [27].

6.4.6. Retrieval of Mars surface physical properties from OMEGA hyperspectral images.

Participants: Stéphane Girard, Alessandro Chiancone.

Joint work with: S. Douté from Laboratoire de Planétologie de Grenoble, J. Chanussot (Gipsa-lab and Grenoble-INP) and J. Saracco (Univ. Bordeaux).

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 (ill-conditioned problems) [57]. We have also defined an adaptive version of the method which is able to deal with block-wise evolving data streams [15].

6.4.7. High-dimensional change-point detection with sparse alternatives

Participant: Farida Enikeeva.

Joint work with: Zaid Harchaoui from LEAR team Inria Grenoble

The change-point problem is a classical problem of statistics that arises in various applications as signal processing, bioinformatics, financial market analysis. The goal of change-point problems is to make an inference about the moment of a change in the distribution of the observed data. We consider the problem of detection of a simultaneous change in mean in a sequence of Gaussian vectors.

The state-of-the-art approach to the change-point detection/estimation is based on the assumption of growing number of observations and fixed dimension of the signal. We work in high-dimensional setting assuming that the vector dimension tends to infinity and the length of the sequence grows slower than the dimension of the signal. Assuming that the change occurs only in a subset of the vector components of unknown cardinality we can reduce our problem to the problem of testing non-zero components in a sequence of sparse Gaussian vectors. We construct a testing procedure that is adaptive to the number of components with a change. This testing procedure is based on combination of two chi-squared type test statistics. This combined test provides an optimal performance of the test both in the cases of high and moderate sparsity. We obtain the detection boundary of the test and show its rate-optimality in minimax sense.

The results of the paper [59] were presented at

- NIPS 2013, Workshop on Modern Nonparametric Methods in Machine Learning (Dec. 2013)
- Conference on Structural Inference in Statistics, Potsdam, Germany (Sept. 2013)

6.4.8. Yield Improvement by the Redundancy Method for Component Calibration

Participant: Farida Enikeeva.

Joint work with: Dominique Morche (CEA-LETI) and Alp Oguz (CEA-LETI)

This work [23] was done in the framework of the Optimyst II project of MINALOGIC in collaboration with CEA-LETI and LJK-UJF. In this project we explore the benefits of the redundant channels methodology for the calibration of electronic components.

The demand for high data rate in communication puts stringent requirements on components' dynamic range. However, the extreme size reduction in advanced technology results inadvertently in increased process variability, which inherently limits the performances. The redundancy approach is based on the idea of dividing an elementary component (capacitor, resistor, transistor) into several subsets and then choosing an optimal combination of such subsets to provide the production of a component with very precise characteristics. For several years, the redundancy method has been identified as complementary to digital calibration to improve the performances. On practice, it is hard for a designer to select an optimal number of redundant components to provide the desired production yield and to minimize the area occupied by the components. The usual way to solve this problem is to resort to statistical simulations which are time consuming and sometimes misleading. We propose a normal approximation of the yield in order to estimate the number of redundant components needed to provide a minimal area occupied by the components.

7. Partnerships and Cooperations

7.1. Regional Initiatives

- MISTIS is involved in three regional initiatives: PEPS (funded by CNRS and the PRES of Grenoble), AGIR (funded by Université Grenoble 1 and Grenoble-INP) and the MOTU project (funded by UPMF). The first two projects focus on the modelling of the extreme risk and its application in social science. The partners include the LTHE (Laboratoire d'étude des Transferts en Hydrologie et Environnement) and the 3S-R lab (Sols, Solides, Structures - Risques). The third project focuses on the use of statistical techniques for transportation data analysis and involves the GAEL laboratory (Grenoble Applied Economics Laboratory).
- 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.
- S. Girard is at the head of the probability and statistics department of the LJK since september 2012.

7.2. National Initiatives

7.2.1. Competitvity Clusters

MISTIS was a partner in a three-year MINALOGIC project (I-VP for Intuitive Vision Programming) supported by the French Government. The project was 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 involved were the CMM (Centre de Morphologie Mathématiques) in Fontainebleau, and Pige Electronique in Bourg-Les-Valence. The overall goal was to exploit statistical and image processing techniques more intensively to improve defect detection capability and programming time based on existing AOI principles so as to eventually reach a reliable defect detection with virtually zero programming skills and efforts. The final review of this project was held in March 2013 with live demos of our tools at VIT.

7.2.2. Inria project HEROES

The 2-year Inria ARC project AINSI (2011-12) coordinated by F. Forbes (<http://thalie.ujf-grenoble.fr/ainsi>) was followed with the same partners by a project entitled HEROES. HEROES stands for "HEmodynamics-infoRmed atlas of brain functiOnal and vascular territoriES from multimodal MR images". The goal, based on ASL and BOLD fMRI and advanced models, is to 1) provide individual brain maps of hemodynamic characteristics useful as biomarkers and 2) extend the use of functional MRI (BOLD or ASL) in the clinic through an improved characterization of the impact of vascular alterations under pathological conditions. The partners involved are Visages and Panama teams from Inria in Rennes and Parietal in Saclay, the INSERM Unit U594 (Grenoble Institute of Neuroscience) and CEA NeuroSpin.

7.3. European Initiatives

7.3.1. FP7 Projects

7.3.1.1. HUMAVIPS

Title: Humanoids with audiovisual skills in populated spaces

Type: COOPERATION (ICT)

Defi: Cognitive Systems and Robotics

Instrument: Specific Targeted Research Project (STREP)

Duration: February 2010 - January 2013

Coordinator: Inria (France)

Others partners: CTU Prague (Czech Republic), University of Bielefeld (Germany), IDIAP (Switzerland), Aldebaran Robotics (France)

See also: <http://humavips.inrialpes.fr>

Abstract: Humanoids expected to collaborate with people should be able to interact with them in the most natural way. This involves significant perceptual and interactive skills, operating in a coordinated fashion. Consider a social gathering scenario where a humanoid is expected to possess certain social skills. It should be able to analyze a populated space, to localize people, and to determine whether they are looking at the robot and are speaking to it. Humans appear to solve these tasks routinely by integrating the often complementary information provided by multi-sensory data processing, from 3D object positioning and sound-source localization to gesture recognition. Understanding the world from unrestricted sensorial data, recognizing people's intentions and behaving like them are extremely challenging problems. The objective of HUMAVIPS has been to endow humanoid robots with audiovisual (AV) abilities: exploration, recognition, and interaction, such that they exhibit adequate behavior when dealing with a group of people. Developed research and technological developments have emphasized the role played by multimodal perception within principled models of human-robot interaction and of humanoid behavior. An adequate architecture has implemented auditory and visual skills onto a fully programmable humanoid robot (the consumer robot NAO). A free and open-source software platform has been developed to foster dissemination and to ensure exploitation of the outcomes of HUMAVIPS beyond its lifetime.

7.4. International Initiatives

7.4.1. Inria International Partners

7.4.1.1. Informal International Partners

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

- Emma Holian and John Hinde from National University of Ireland, Galway, Ireland.
- K. Qin and D. Wraith from RMIT and Centre for Epidemiology and Biostatistics University in Melbourne, Australia.
- E. Deme and S. Sylla from Saint Louis university and IRD in Saint Louis, 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.

7.5. International Research Visitors

7.5.1. Visits of International Scientists

- Alexander Nazin (Russian Academy of Sciences, Russia) has been an invited researcher in the MISTIS team to work with Stéphane Girard and Anatoli Ioudistki (Université Grenoble 1).
- El Hadji Deme (Université Gaston Berger, Sénégal) has been hosted by the MISTIS team for two months. His stay has been partially funded by the Ibni Oumar Mahamat Saleh price.

7.5.1.1. Internships

Jennifer Sloboda (Master, from May 2013 until Aug 2013)

Subject: Physiologically-inspired Bayesian analysis of BOLD and ASL fMRI data

Institution: University of Michigan, Ann Arbor (United States)

Aina Frau-Pascual (Master, from May 2013 until Aug 2013)

Subject: Hemodynamically informed parcellation of cerebral fMRI data

Institution: University Paris Sud

Pham Van Trung (Master, from Apr 2013 until Sep 2013)

Subject: Implémentation et paquetage d'un modèle statistique des valeurs extrêmes.
Institution: Hanoi, Vietnam.

Seydou-Nourou Sylla (PhD, from October 2013 to December 2013)

Subject: Classification for medical data
Institution: Université Gaston Berger (Sénégal)

8. Dissemination

8.1. Scientific Animation

Editorial activities

- Stéphane Girard is Associate Editor of the *Statistics and Computing* journal since 2012. He has been also an invited editor for a special issue of the *Journal de la Société Française de Statistique* dedicated to extreme-value analysis.

Workshops and summer schools

- Florence Forbes and Stéphane Girard co-organized the summer school “*Méthodes et applications de la régression en astrophysique*”, Annecy, <http://ecastrostat2013.sciencesconf.org/>.
- Florence Forbes and Stéphane Girard co-organized the workshop “*Géométrie Aléatoire et ses Applications*”, Grenoble, <http://gdr-geostoch.math.cnrs.fr/index.html>.
- Stéphane Girard organized the workshop “*Copulas and extremes*”, Grenoble, <http://mistis.inrialpes.fr//workshop-copulas-extremes.html>.
- Marie José Martinez, Jean Baptiste Durand, Florence Forbes in collaboration with Iragael Joly (Grenoble Applied Economics Laboratory) organized the workshop "Statistics, Activities and Transportation" in Grenoble <http://mistis.inrialpes.fr/workshop-statistique-transport.html> as part of the MOTU project (2013-14).

Societies and Networks

- F. Forbes is part of an INRA (French National Institute for Agricultural Research) Network (AIGM, <http://carlit.toulouse.inra.fr/AIGM>) on Algorithmic issues for inference in graphical models.
- F. Forbes and S. Girard were elected as members of the bureau of the “Analyse d’images, quantification, et statistique” group in the Société Française de Statistique (SFdS).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence (IUT): Marie-José Martinez , *Statistics*, 192 ETD, L1 to L3 levels, université Grenoble 2, France.

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

Licence (IUT) : Gildas Mazo, *Mathematics and C language*, 128h, L1 level, université Grenoble 1, France.

Master: Farida Enikeeva, *Statistics*, 96 ETD, M1 level, Ensimag Grenoble INP, France.

Licence: Christine Bakhous, *Mathematics and Statistics*, 64 ETD, L1 level, université Grenoble 1, France.

Licence: Jonathan El Methni, *Mathematics and Statistics*, 64 ETD, L1 level, université Grenoble 1, France.

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

Master : Florence Forbes, *Mixture models and EM algorithm*, 12h, M2 level, UFR IM2A, université Grenoble 1, France.

M.-J. Martinez is faculty members at Univ. Pierre Mendès France, Grenoble II.

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

F. Enikeeva is on a half-time ATER position at Ensimag, Grenoble INP.

C. Bakhous and J. El Methni were both moniteur at University Joseph Fourier.

8.2.2. Supervision

PhD: Christine Bakhous, "*Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par IRM fonctionnelle*", Université Grenoble 1, defended on December 2013. Supervision : Florence Forbes & Michel Dojat (GIN).

PhD : Jonathan El-methni, "*Différentes contributions à l'estimation de quantiles extrêmes*" Université Grenoble 1, defended on october 2013. Supervision : Stéphane Girard & Laurent Gardes (Université de Strasbourg).

PhD : El-hadji Deme, "*Quelques contributions à la théorie univariée des valeurs extrêmes. Estimation des mesures de risque actuariel pour des pertes à queues lourdes*", Université Gaston Berger, Sénégal, defended on june 2013. Supervision : Stéphane Girard & Gane Samb Lo (Université Gaston Berger, Sénégal)

PhD in progress: Aina Frau-Pascual, "*Statistical Models for the coupling of ASL and BOLD Magnetic Resonance modalities to study brain function and disease*", Université Grenoble 1, started in october 2013. Supervision : Florence Forbes & Philippe Ciuciu (Parietal, NeuroSpin).

PhD in progress : Alessandro Chiancone "*Sequential dimension reduction*", Université Grenoble 1, started in october 2013. Supervision : Stéphane Girard & Jocelyn Chanussot (Gypsa-lab, 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*", Université Gaston Berger, Sénégal, started in october 2012. Supervision : Stéphane Girard & Abdou Ka Diongue (Université Gaston Berger, Sénégal).

PhD in progress : Gildas Mazo, "*Estimation de quantiles extrêmes spatiaux à partir de données environnementales*", Université Grenoble 1, started in october 2011. Supervision : Florence Forbes & Stéphane Girard.

8.2.3. Juries

Stéphane Girard has been involved in the following PhD committees:

- Yousri Henchiri, "*Support Vector Machine (SVM) pour l'analyse de données fonctionnelles*", Université Montpellier 2.
- François Portier "*Réduction de la dimension en régression*", Université Rennes 1.
- Smriti Joshi, "*Consommation statique dans les circuits numériques en CMOS 32nm: Analyse et méthodologie pour une estimation statistique au niveau porte*", Université Grenoble.

Florence Forbes has been involved in the PhD committees of:

- Xavier Alameda-Pineda, Egocentric audio-visual scene analysis: a machine learning and signal processions approach, University Grenoble 1.
- Antoine Deleforge, Acoustic Space Mapping: a machine learning approach to sound source separation and localization, University Grenoble 1
- Mohamad Belouni, Plans d'expérience optimaux en régression appliquée à la pharmacocinétique, University Grenoble 1.
- Solveig Badillo, Etude de la variabilité hémodynamique chez l'enfant et l'adulte sains en IRMf, University Paris Sud
- Virgile Fritsch, High-dimensional statistical methods for inter-subjects studies in neuroimaging, University Paris Sud

Since September 2009, F. Forbes is head of the committee in charge of examining post-doctoral candidates at Inria Grenoble Rhône-Alpes ("Comité des Emplois Scientifiques").

Florence Forbes is a member of the INRA committee (CSS MBIA) in charge of evaluating INRA researchers once a year in the MBIA dept of INRA.

Florence Forbes was a member of:

- the AERES committee in charge of evaluating the AgroParisTech unit.
- the committee for selecting a new professor at University Grenoble 1.
- the LJK committee for attributing the first Jean Kuntzman award.

9. 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 Analysis 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] G. CELEUX, S. CHRÉTIEN, F. FORBES, A. MKHADRI. *A Component-wise EM Algorithm for Mixtures*, in "Journal of Computational and Graphical Statistics", 2001, vol. 10, pp. 699–712
- [6] G. CELEUX, F. FORBES, N. PEYRARD. *EM procedures using mean field-like approximations for Markov model-based image segmentation*, in "Pattern Recognition", 2003, vol. 36, n^o 1, pp. 131-144
- [7] F. FORBES, G. FORT. *Combining Monte Carlo and Mean field like methods for inference in hidden Markov Random Fields*, in "IEEE trans. PAMI", 2007, vol. 16, n^o 3, pp. 824-837
- [8] F. FORBES, N. PEYRARD. *Hidden Markov Random Field Model Selection Criteria based on Mean Field-like Approximations*, in "IEEE trans. PAMI", August 2003, vol. 25(9), pp. 1089–1101
- [9] S. GIRARD. *A Hill type estimate of the Weibull tail-coefficient*, in "Communication in Statistics - Theory and Methods", 2004, vol. 33, n^o 2, pp. 205–234
- [10] S. GIRARD, P. JACOB. *Extreme values and Haar series estimates of point process boundaries*, in "Scandinavian Journal of Statistics", 2003, vol. 30, n^o 2, pp. 369–384

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [11] C. BAKHOUS. , *Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par IRM fonctionnelle*, Université de Grenoble, December 2013, <http://hal.inria.fr/tel-00933426>
- [12] J. EL METHNI. , *Contributions à l'estimation de quantiles extrêmes. Applications à des données environnementales*, Université de Grenoble, October 2013, <http://hal.inria.fr/tel-00924293>

Articles in International Peer-Reviewed Journals

- [13] S. BADILLO, T. VINCENT, P. CIUCIU. *Group-level impacts of within- and between-subject hemodynamic variability in fMRI*, in "NeuroImage", June 2013, vol. 82C, pp. 433-448 [DOI : 10.1016/J.NEUROIMAGE.2013.05.100], <http://hal.inria.fr/hal-00854481>
- [14] 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>
- [15] M. CHAVENT, S. GIRARD, V. KUENTZ, B. LIQUET, T. M. N. NGUYEN, J. SARACCO. *A sliced inverse regression approach for data stream*, in "Computational Statistics", 2014, vol. to appear, <http://hal.inria.fr/hal-00688609>
- [16] A. DAOUIA, L. GARDES, S. GIRARD. *On kernel smoothing for extremal quantile regression*, in "Bernoulli", November 2013, vol. 19, n^o 5B, pp. 2557-2589 [DOI : 10.3150/12-BEJ466], <http://hal.inria.fr/hal-00630726>
- [17] A. DAOUIA, S. GIRARD, A. GUILLOU. *A Gamma-moment approach to monotonic boundaries estimation: with applications in econometric and nuclear fields*, in "Journal of Econometrics", February 2014, vol. 178, n^o 2, pp. 727-740 [DOI : 10.1016/J.JECONOM.2013.10.013], <http://hal.inria.fr/hal-00737732>
- [18] E. H. DEME, L. GARDES, S. GIRARD. *On the estimation of the second order parameter for heavy-tailed distributions*, in "REVSTAT - Statistical Journal", November 2013, vol. 11, n^o 3, pp. 277-299, <http://hal.inria.fr/hal-00634573>
- [19] E. H. DEME, S. GIRARD, A. GUILLOU. *Reduced-bias estimator of the Proportional Hazard Premium for heavy-tailed distributions*, in "Insurance: Mathematics and Economics", March 2013, vol. 52, n^o 3, pp. 550-559 [DOI : 10.1016/J.INSMATHECO.2013.03.010], <http://hal.inria.fr/hal-00763978>
- [20] J.-B. DURAND, S. GIRARD, V. CIRIZA, L. DONINI. *Optimization of power consumption and user impact based on point process modeling of the request sequence*, in "Journal of the Royal Statistical Society Series C", March 2013, vol. 62, n^o 2, pp. 151-165 [DOI : 10.1111/J.1467-9876.2012.01053.X], <http://hal.inria.fr/hal-00412509>
- [21] J.-B. DURAND, B. GUITTON, J. PEYHARDI, Y. HOLTZ, Y. GUÉDON, C. TROTTIER, E. COSTES. *New insights for estimating the genetic value of F1 apple progenies for irregular bearing during the first years of tree production*, in "Journal of Experimental Botany", November 2013, vol. 64, n^o 16, pp. 5099-5113 [DOI : 10.1093/JXB/ERT297], <http://hal.inria.fr/hal-00845353>

- [22] J. EL METHNI, L. GARDES, S. GIRARD. *Nonparametric estimation of extreme risk measures from conditional heavy-tailed distributions*, in "Scandinavian Journal of Statistics", 2014, to appear, <http://hal.inria.fr/hal-00830647>
- [23] F. ENIKEEVA, D. MORCHE, A. OGUZ. *Yield Improvement by the Redundancy Method for Component Calibration*, in "Electronics Letters", June 2013, vol. 49, n^o 13, pp. 824-826 [DOI : 10.1049/EL.2012.4170], <http://hal.inria.fr/hal-00932560>
- [24] F. FORBES, M. CHARRAS-GARRIDO, L. AZIZI, S. DOYLE, D. ABRIAL. *Spatial risk mapping for rare disease with hidden Markov fields and variational EM*, in "Annals Of Applied Statistics", June 2013, vol. 7, n^o 2, pp. 1192-1216 [DOI : 10.1214/13-AOAS629], <http://hal.inria.fr/hal-00839184>
- [25] 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", 2013, To appear [DOI : 10.1007/s11222-013-9414-4], <http://hal.inria.fr/hal-00823451>
- [26] S. GIRARD, A. GUILLOU, G. STUPFLER. *Frontier estimation with kernel regression on high order moments*, in "Journal of Multivariate Analysis", April 2013, vol. 116, pp. 172-189 [DOI : 10.1016/J.JMVA.2012.12.001], <http://hal.inria.fr/hal-00499369>
- [27] S. GIRARD, A. GUILLOU, G. STUPFLER. *Uniform strong consistency of a frontier estimator using kernel regression on high order moments*, in "ESAIM: Probability and Statistics", 2013, to appear [DOI : 10.1051/PS/2013050], <http://hal.inria.fr/hal-00764425>
- [28] M.-J. MARTINEZ, E. HOLIAN. *An alternative estimation approach for the heterogeneity linear mixed model*, in "Communications in Statistics - Simulation and Computation", September 2013, (to appear) [DOI : 10.1080/03610918.2012.762389], <http://hal.inria.fr/hal-00926620>

Articles in National Peer-Reviewed Journals

- [29] L. GARDES, S. GIRARD. *Estimation de quantiles extrêmes pour les lois à queue de type Weibull : une synthèse bibliographique*, in "Journal de la Société Française de Statistique", 2013, vol. 154, n^o 2, pp. 98-118, <http://hal.inria.fr/hal-00799249>

International Conferences with Proceedings

- [30] 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, 2013, pp. 1512-1515 [DOI : 10.1109/ISBI.2013.6556822], <http://hal.inria.fr/hal-00854624>
- [31] C. BAKHOUS, F. FORBES, T. VINCENT, M. DOJAT, P. CIUCIU. *Variational Variable Selection To Assess Experimental Condition Relevance In Event-Related fMRI*, in "ISBI 2013 - 10th IEEE International Symposium on Biomedical Imaging", San Francisco, United States, IEEE, April 2013, pp. 1508-1511 [DOI : 10.1109/ISBI.2013.6556821], <http://hal.inria.fr/hal-00859391>
- [32] A. DAMBREVILLE, P. FERNIQUE, C. PRADAL, P.-E. LAURI, F. NORMAND, Y. GUÉDON, J.-B. DURAND. *Deciphering mango tree asynchronisms using Markov tree and probabilistic graphical models*, in "FSPM2013 - 7th International Workshop on Functional-Structural Plant Models", Saariselkä, Finland, R. SIEVÄNEN, E. NIKINMAA, C. GODIN, A. LINTUNEN, P. NYGREN (editors), 2013, pp. 210-212, ISBN 978-951-651-408-9, <http://hal.inria.fr/hal-00847614>

- [33] A. DAOUIA, L. GARDES, S. GIRARD. *Nonparametric extremal quantile regression*, in "EVA 2013 - 8th International Conference on Extreme Value Analysis", Shanghai, China, 2013, <http://hal.inria.fr/hal-00851124>
- [34] A. DAOUIA, S. GIRARD, A. GUILLOU. *A gamma-moment approach to monotonic boundaries estimation: with applications in econometric and nuclear fields*, in "EVA 2013 - 8th International Conference on Extreme Value Analysis", Shanghai, China, 2013, <http://hal.inria.fr/hal-00851125>
- [35] A. DELEFORGE, F. FORBES, R. HORAUD. *Variational EM for Binaural Sound-Source Separation and Localization*, in "ICASSP 2013 - 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, IEEE, 2013, pp. 76-80 [DOI : 10.1109/ICASSP.2013.6637612], <http://hal.inria.fr/hal-00823453>
- [36] E. H. DEME, S. GIRARD, A. GUILLOU. *Reduced-biased estimators of the Conditional Tail Expectation for heavy-tailed distributions*, in "Mathematical Statistics and Limit Theorems, Conference in honor of Prof. Paul Deheuvels", Paris, France, 2013, <http://hal.inria.fr/hal-00837681>
- [37] J.-B. DURAND, B. GUITTON, J. PEYHARDI, Y. HOLTZ, Y. GUÉDON, C. TROTTIER, E. COSTES. *Estimating the genetic value of F1 apple progenies for irregular bearing during first years of production*, in "FSPM2013 - 7th International Workshop on Functional-Structural Plant Models", Saariselkä, Finland, R. SIEVÄNEN, E. NIKINMAA, C. GODIN, A. LINTUNEN, P. NYGREN (editors), June 2013, pp. 13-15, <http://hal.inria.fr/hal-00847593>
- [38] J. EL METHNI, L. GARDES, S. GIRARD. *Nonparametric estimation of extreme risks from heavy-tailed distributions*, in "Extremes in Vimeiro Today", Vimeiro, Portugal, September 2013, <http://hal.inria.fr/hal-00862759>
- [39] J. EL METHNI, S. GIRARD, L. GARDES. *Estimation of Extreme Risk Measures from Heavy-tailed distributions*, in "EVA 2013 - 8th Conference on Extreme Value Analysis, probabilistic and statistical models and their applications", Shanghai, China, 2013, <http://hal.inria.fr/hal-00845527>
- [40] S. JOSHI, A. LOMBARDOT, M. BELLEVILLE, E. BEIGNE, S. GIRARD. *A gate level methodology for efficient statistical leakage estimation in complex 32nm circuits*, in "DATE 2013 - Design, Automation and Test in Europe", Grenoble, France, 2013, <http://hal.inria.fr/hal-00805478>
- [41] *Best Paper*
V. KHALIDOV, F. FORBES, R. HORAUD. *Alignment of Binocular-Binaural Data Using a Moving Audio-Visual Target*, in "MMSP 2013 - IEEE International Workshop on Multimedia Signal Processing", Pula (Sardinia), Italy, IEEE, September 2013, pp. 242-247, Best Paper Award [DOI : 10.1109/MMSP.2013.6659295], <http://hal.inria.fr/hal-00861482>.
- [42] G. MAZO, S. GIRARD, F. FORBES. *A parsimonious multivariate copula for tail dependence modeling*, in "EVT 2013 - Extremes in Vimeiro Today", Vimeiro, Portugal, September 2013, <http://hal.inria.fr/hal-00863540>
- [43] T. VINCENT, F. FORBES, P. CIUCIU. *Bayesian BOLD and perfusion source separation and deconvolution from functional ASL imaging*, in "ICASSP 2013 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, IEEE, May 2013, pp. 1003-1007 [DOI : 10.1109/ICASSP.2013.6637800], <http://hal.inria.fr/hal-00859373>

- [44] T. VINCENT, J. WARNKING, M. VILLIEN, A. KRAINIK, P. CIUCIU, F. FORBES. *Bayesian Joint Detection-Estimation of cerebral vasoreactivity from ASL fMRI data*, in "MICCAI 2013 - 16th International Conference on Medical Image Computing and Computer Assisted Intervention", Nagoya, Japan, K. MORI, I. SAKUMA, Y. SATO, C. BARILLOT, N. NAVAB (editors), Lecture Notes in Computer Science, Springer, June 2013, vol. 8150, pp. 616-623 [DOI : 10.1007/978-3-642-40763-5_76], <http://hal.inria.fr/hal-00854437>

National Conferences with Proceedings

- [45] C. BAKHOUS, F. FORBES, F. ENIKEEVA, T. VINCENT, M. DOJAT, P. CIUCIU. *Analyse parcimonieuse des données d'IRM fonctionnelle dans un cadre bayésien variationnel*, in "45èmes Journées de Statistique", Toulouse, France, Société Française de Statistique, May 2013, <http://hal.inria.fr/hal-00933701>
- [46] J. EL METHNI, L. GARDES, S. GIRARD. *Estimation de mesures de risque extrêmes*, in "45èmes Journées de Statistique", Toulouse, France, 2013, organisées par la Société Française de Statistique, <http://hal.inria.fr/hal-00829625>
- [47] J. EL METHNI, S. GIRARD, L. GARDES. *Estimation de mesures de risque pour des pluies extrêmes dans la région Cévennes-Vivarais*, in "Evénements extrêmes d'inondation : de l'étude de l'aléa à la gestion du risque pour les ouvrages hydrauliques", Lyon, France, 2013, <http://hal.inria.fr/hal-00905192>
- [48] A. STUDENY, F. FORBES, E. COISSAC, A. VIARI, C. MERCIER, L. ZINGER, A. BONIN, F. BOYER, P. TABERLET. *Spatial modelling of plant diversity from high-throughput environmental dna sequence data*, in "45èmes Journées de Statistique", Toulouse, France, Société Française de Statistique, May 2013, <http://hal.inria.fr/hal-00933699>

Conferences without Proceedings

- [49] E. H. DEME, L. GARDES, S. GIRARD. *Estimation of the second order parameter for heavy-tailed distributions*, in "ERCIM 2013 - 6th International Conference of the ERCIM WG on Computational and Methodological Statistics", London, United Kingdom, December 2013, <http://hal.inria.fr/hal-00915698>
- [50] J.-B. DURAND, P. FERNIQUE. *Approche graphique pour la modélisation statistique de la dépendance entre activités journalières*, in "Workshop "Statistique, Transport et Activités"", Grenoble, France, November 2013, <http://mistis.inrialpes.fr/workshop-statistique-transport.html>, <http://hal.inria.fr/hal-00919372>
- [51] P. FERNIQUE, J.-B. DURAND, Y. GUÉDON. *Parametric Modelling of Multivariate Count Data Using Probabilistic Graphical Models*, in "AIGM13 - 3rd Workshop on Algorithmic issues for Inference in Graphical Models", Paris, France, September 2013, <http://hal.inria.fr/hal-00918597>
- [52] M.-J. MARTINEZ, E. HOLIAN. *An alternative estimation approach to fit a heterogeneity linear mixed model*, in "ERCIM 2013 - 6th International Conference of the ERCIM Working Group on Computational and Methodological Statistics", Londres, United Kingdom, December 2013, <http://hal.inria.fr/hal-00926627>
- [53] G. MAZO, S. GIRARD, F. FORBES. *A copula to handle tail dependence in high dimension*, in "ERCIM 2013 - 6th International Conference of the ERCIM WG on Computational and Methodological Statistics", London, United Kingdom, December 2013, <http://hal.inria.fr/hal-00915690>

Scientific Books (or Scientific Book chapters)

- [54] P. CIUCIU, F. FORBES, T. VINCENT, L. CHAARI. *Détection-estimation conjointe en IRM fonctionnelle*, in "Méthodes d'inversion appliquées au traitement du signal et de l'image", J.-F. GIOVANNELLI, J. IDIER (editors), Hermès, November 2013, <http://hal.inria.fr/hal-00823455>

Research Reports

- [55] A. DELEFORGE, F. FORBES, R. HORAUD. , *High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables*, December 2013, <http://hal.inria.fr/hal-00863468>

Other Publications

- [56] C. AMBLARD, S. GIRARD, L. MENNETEAU. , *Bivariate copulas defined from matrices*, October 2013, <http://hal.inria.fr/hal-00875303>

- [57] R. COUDRET, S. GIRARD, J. SARACCO. , *A new sliced inverse regression method for multivariate response*, 2013, <http://hal.inria.fr/hal-00714981>

- [58] E. H. DEME, S. GIRARD, A. GUILLOU. , *Reduced-bias estimator of the Conditional Tail Expectation of heavy-tailed distributions*, September 2013, <http://hal.inria.fr/hal-00823260>

- [59] F. ENIKEEVA, Z. HARCHAOUI. , *High-dimensional change-point detection with sparse alternatives*, 2013, <http://hal.inria.fr/hal-00933185>

- [60] S. GIRARD, G. STUPFLER. , *Intriguing properties of extreme geometric quantiles*, September 2013, <http://hal.inria.fr/hal-00865767>

- [61] G. MAZO, S. GIRARD, F. FORBES. , *A class of high dimensional copulas based on products of bivariate copulas*, 2013, <http://hal.inria.fr/hal-00910775>

- [62] B. MENZE, A. JAKAB, S. BAUER, J. KALPATHY-CRAMER, K. FARAHANI, J. KIRBY, Y. BURREN, N. PORZ, J. SLOTBOOM, R. WIEST, L. LANCZI, E. GERSTNER, M.-A. WEBER, T. ARBEL, B. AVANTS, N. AYACHE, P. BUENDIA, L. COLLINS, N. CORDIER, J. CORSO, A. CRIMINISI, T. DAS, H. DELINGETTE, C. DEMIRALP, C. DURST, M. DOJAT, S. DOYLE, J. FESTA, F. FORBES, E. GEREMIA, B. GLOCKER, P. GOLLAND, X. GUO, A. HAMAMCI, K. IFTEKHARUDDIN, R. JENA, N. JOHN, E. KONUKOGLU, D. LASHKARI, J. ANTONIO MARIZ, R. MEIER, S. PEREIRA, D. PRECUP, S. J. PRICE, T. RIKLIN-RAVIV, S. REZA, M. RYAN, L. SCHWARTZ, H.-C. SHIN, J. SHOTTON, C. SILVA, N. SOUSA, N. SUBBANNA, G. SZEKELY, T. TAYLOR, O. THOMAS, N. TUSTISON, G. UNAL, F. VASSEUR, M. WINTERMARK, D. HYE YE, L. ZHAO, B. ZHAO, D. ZIKIC, M. PRASTAWA, M. REYES, K. VAN LEEMPUT. , *The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)*, January 2014, <http://hal.inria.fr/hal-00935640>

References in notes

- [63] C. BAKHOUS, F. FORBES, T. VINCENT, L. CHAARI, M. DOJAT, P. CIUCIU. *Adaptive experimental condition selection in event-related fMRI*, in "ISBI 2012 - IEEE International Symposium on Biomedical Imaging", Barcelone, Spain, IEEE, May 2012, pp. 1755-1758 [DOI : 10.1109/ISBI.2012.6235920], <http://hal.inria.fr/cea-00710489>

- [64] 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>

-
- [65] P. EMBRECHTS, C. KLÜPPELBERG, T. MIKOSH. , *Modelling Extremal Events*, Applications of Mathematics, Springer-Verlag, 1997, vol. 33
- [66] F. FERRATY, P. VIEU. , *Nonparametric Functional Data Analysis: Theory and Practice*, Springer Series in Statistics, Springer, 2006
- [67] 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
- [68] Y. GOEGBEUR, J. BEIRLANT, T. DE WET. *Kernel estimators for the second order parameter in extreme value statistics*, in "Journal of Statistical Planning and Inference", 2010, vol. 140, n^o 9, pp. 2632–2652
- [69] M. GOMES, L. DE HAAN, L. PENG. *Semi-parametric Estimation of the Second Order Parameter in Statistics of Extremes*, in "Extremes", 2002, vol. 5, n^o 4, pp. 387–414
- [70] D. KOLLER, N. FRIEDMAN. , *Probabilistic graphical models: principles and techniques*, MIT press, 2009
- [71] K. LI. *Sliced inverse regression for dimension reduction*, in "Journal of the American Statistical Association", 1991, vol. 86, pp. 316–327
- [72] R. NELSEN. , *An introduction to copulas*, Lecture Notes in Statistics, Springer-VerlagNew-York, 1999, vol. 139