INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team MISTIS

Modelling and Inference of Complex and Structured Stochastic Systems

Rhône-Alpes

THEME COG

Activity Report

2004

Table of contents

1. Team

Team leader
Florence Forbes [Research scientist, Inria]

Research scientist
Paulo Gonçalvès [Research scientist, Inria]
Christian Lavergne [Faculty member, Prof. Univ. Montpellier]

Project technical staff
Grégory Noulin [until June 2004]

Ph. D. student
Juliette Blanchet [MENRT since October 2004]
Charles Bouveyron [MENRT]
Julien Jacques [Inria until September 2004]
Matthieu Vignes [AC]
Mohamed Saidane [Teaching Assistant, Tunis]

Research scientist (partner)
Henri Bertholon [Faculty member, CNAM, Paris]
Gersende Fort [Research scientist, CNRS, Grenoble]
Laurent Gardes [Faculty member, UPMF, Grenoble, since October 2004]
Stéphane Girard [Faculty member, UJF, Grenoble]

Administrative assistant
Élodie Toihein

2. Overall Objectives

The team MISTIS aims at developing statistical methods for dealing with complex problems or data. Our applications consist mainly of image processing and spatial data problems with some applications in 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, Markovian models, and more generally hidden structure models identified by stochastic algorithms on one hand, and semi and non-parametric methods on the other hand.

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

Semi and non-parametric methods are relevant and useful when no appropriate parametric model exist 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).

3. Scientific Foundations

3.1. Mixture models

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

Participants: Juliette Blanchet, Florence Forbes, Gersende Fort, Matthieu Vignes.

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 divide into observed data $y = y_1, ..., y_n$ and unobserved or missing data $z = z_1, ..., z_n$. The missing data z_i represents for instance the memberships to one of a set of K alternative categories. The distribution of an observed y_i can be written as a finite mixture of distributions,

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

These models are interesting in that they may point out an 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 parameters estimation but also values for missing data.

Mixture models correspond to independent z_i's. They are more and more used in statistical pattern recognition. They allow a formal (model-based) approach to (unsupervised) clustering.

3.2. Markovian models

Keywords: *Bayesian inference*, *EM algorithm*, *hidden Markov field*, *image analysis*, *missing data*, *mixture of distributions*, *selection and combination of models*, *statistical pattern recognition*, *stochastic algorithms*.

Participants: Juliette Blanchet, Florence Forbes, Gersende Fort, Paulo Gonçalvès, Christian Lavergne, Mohammed Saidane, Matthieu Vignes.

Hidden Markov chains or hidden Markov fields correspond to cases where the z_i's are distributed according to a Markov chain or a Markov field. These models are widely used in signal processing (speech recognition, genome sequence analysis) and in image processing (remote sensing, MRI, etc.). Markovian models are part of *graphical models*. In these models, the variable organization can be represented by a graph where the nodes represent the variables and the edges the statistical dependencies between the variables. 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, we refer to these models as Hidden Markov Models (HMM). Such models are very flexible in practice and can naturally account for the phenomena to be studied. They 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 but we propose to use variational approximations for estimation and model selection when exact calculations are intractable. Many experiments have to be carried out to assess the approximations quality and the associated estimation methods performance before addressing theoretical properties such as convergence and speed results.

3.3. Functional Inference, semi and non parametric methods

Keywords: *boundary estimation*, *extremes*, *non parametric*, *scaling laws*, *singularity spectra*, *wavelets*.

Participants: Charles Bouveyron, Laurent Gardes, Stéphane Girard, Paulo Goncalvès.

We also consider methods which do not assume a parametric model. Such methods are used for instance to study distribution tails without introducing a parametric model on the data: this is part of the *extreme values theory*. Similarly, the grey-levels surface in an image cannot usually be described through a simple mathematical equation. Projection methods are then a way to decompose the unknown signal or image 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 *boundary estimation*. As regards wavelets, our goal is to propose wavelet based estimators aimed at characterizing and analyzing scaling laws structures of processes or systems. The compression/dilation operator, at the core of wavelet analysis, allows to identify complex scale organizations, such as 1/f type processes (e.g.mono-fractals), high order statistics governed by power laws (e.g. multi-fractals), or more generally cascade type constructions of measures and processes.

4. Application Domains

4.1. Image Analysis

Participants: Juliette Blanchet, Charles Bouveyron, Florence Forbes, Stéphane Girard, Paulo Goncalvès.

As regards applications, several areas of image analysis can be covered using the tools developed in the team. More specifically, we address in collaboration with Team Lear, Inria Rhone-Alpes, issues about object and class recognition and about the extraction of visual information from large image data bases.

Other applications in medical imaging are natural. We worked more specifically on MRI data.

We also consider other statistical 2D fields coming from other domains such as the turbulent velocity fields or the representations of 1D signals on a time-frequency plane.

4.2. Biology and Medicine

Participants: Florence Forbes, Christian Lavergne, Matthieu Vignes.

A second domain of applications concerns biomedical statistics and molecular biology. We consider the use of missing data models in epidemiology. We also investigate statistical tools for the analysis of bacterial genomes beyond gene detection.

4.3. Reliability

Participants: Henri Bertholon, Julien Jacques, Christian Lavergne.

Reliability and industrial lifetime analysis are applications developed essentially through collaborations with the EDF research department and the LCFR laboratory of CEA / Cadarache.

5. Software

5.1. MixMod (Mixture Modelling) freeware

Participant: Grégory Noulin.

Joint work with Christophe Biernacki and Florent Langrognet (Université de Franche-Comté) and Gérard Govaert (Université de Technologie de Compiègne).

MIXMOD (MIXture MODelling) software fits multivariate Gaussian mixtures to a given data set with either a density estimation, a cluster analysis or a discriminant analysis point of view. This software is original in three ways.

- A large variety of algorithms to estimate the mixture parameters are proposed (EM, Classification EM, Stochastic EM) and it is possible to combine them to lead to different strategies to get a sensible maximum of the likelihood function.
- Moreover, 28 different mixture models can be considered according to different assumptions on the component variance matrix eigenvalue decomposition.
- Finally, different information criteria for choosing a parsimonious model, some of them favoring a cluster analysis view point, are included.

Written in C++, MIXMOD is easily interfaced with Scilab and Matlab. It can be downloaded at the following URL: http://www-math.univ-fcomte.fr/mixmod/index.php.

5.2. The Extremes freeware

Participant: Stéphane Girard.

Joint work with Jean Diebolt (CNRS), Myriam Garrido (ENAC, Université Toulouse 3) and Jérôme Ecarnot.

The EXTREMES software is a toolbox dedicated to the modelling of extremal events offering extreme quantile estimation procedures and model selection methods [23]. This software results from a collaboration with EDF R&D. It is also a consequence of the PhD thesis work of Myriam Garrido. The software is written in C++ with a Matlab graphical interface. It can be downloaded at the following URL: http://www.inrialpes.fr/is2/pub/software/EXTREMES/accueil.html. Recently, this software has been used to propose a new goodness-of-fit test to the distribution tail [21].

6. New Results

6.1. Mixture models

6.1.1. Model-based Region Of Interest selection in dynamic breast MRI

Participant: Florence Forbes.

Joint work with Nathalie Peyrard (INRA, Avignon), Chris Fraley and Adrian Raftery (Statistics Department, University of Washington, Seattle).

The project started as a collaboration between the University of Washington Department of Statistics, the University of Washington Breast Imaging Center, and Toshiba America MRI, with the latter two collaborating to acquire the data. In a first version of this work [25] only three patients were analysed. We then had to extend the analysis and analyze data from patients that we had not analyzed for the first version. Changes in some of the participants affiliation, together with human subjects constraints, meant that it was logistically complicated for us to get permission to analyze the new data and then to recover them. In the end we did succeed to extend our analyis to 19 patients, and the results were very good [19].

6.2. Markovian models

6.2.1. Convergence properties of EM like algorithms for inference in Hidden Markov Random Fields

Participants: Florence Forbes, Gersende Fort.

Outside simple cases, the EM algorithm is seldom tractable analytically. In practice, difficulties arise due to the dependence structure in the models and approximations are required. A heuristic solution using mean field approximation principle has been proposed in [31]. Using ideas from this principle, we proposed [3] a class of EM-like algorithms generalizing [31]. The mean field approach consists of neglecting fluctuations from the mean in the environment of each variable. More generally, we talk about mean field-like approximations when the value at node i does not depend on the values at other nodes which are all set to constants (not necessarily the means) independently of the value at node i ([3]). The following computation then reduces to dealing with systems of independent variables, which is much simpler.

This approach is very flexible in that many ways to set the neighboring nodes are possible and lead to as many different algorithms. We investigated some of these choices [3][6] which led to promising procedures. Their behavior is satisfying in practice but no theoretical study as regards convergence properties is available yet.

To investigate such convergence properties, we propose to consider a particular way to set the neighbors which induces the increase of a function of interest. The function is chosen so as to facilitate the the convergence study of the subsequent algorithm. After implementing and assessing the performance of this algorithm in practice, a second step is to consider techniques developped in [26] to link the properties of the algorithm to the other algorithms originally developped in [3].

6.2.2. *Factorial Hidden Markov Models for time series in finance*

Participants: Christian Lavergne, Mohamed Saidane.

The purpose of our work is the development of dynamic factor models for multivariate financial time series, and the incorporation of stochastic volatility components for latent factor processes. The models are direct generalizations of univariate stochastic volatility models, and represent specific varieties of models recently discussed in the growing multivariate stochastic volatility literature.

6.2.3. *Stastistical tools for the analysis of bacterial genomes organisation*

Participants: Florence Forbes, Matthieu Vignes.

We investigated a part of the exploratory analysis of bacterial genomes, beyond gene detection. We aim at detecting relationships among genes based on different kinds of information: nucleotide sequence, gene position, functional annotation,... The ideal goal is to link proximities among genes on the chromosome with genetic mechanisms of the cell. In fact, the cell machinery is thought to be coded inside the genome. We reviewed the main work in progress on the subject in order to suggest an appropriate formalism. We focused on the notion of neighborhood, related to intrinsic properties among entities (genes) considered. Neighborhood must be understood in a broad sense which leads to some specific mathematical tools and processes. Our investigation is based on tools from mixture models and markovian models. We consider various classification methods.

6.2.4. *Markov Random Fields for recognizing textures*

Participants: Florence Forbes, Juliette Blanchet.

We present a new probabilistic framework for recognizing textures in images. Images are described by local affine-invariant descriptors and by spatial relationships between these descriptors. A graph is associated to an image with the nodes representing feature vectors describing image regions and the edges joining spatially related regions. Incorporating information about the spatial organization of the descriptors leads to better recognition results. Current approaches consist in augmenting the data with information coming from the spatial relationships, for instance by using co-occurence statistics, but without modeling explicitly the dependencies between neighboring descriptors. In such approaches the underlying model is one where the descriptors are statistically independent variables. Our claim is that recognition results can be further improved by considering that descriptors are statistically dependent. We propose to introduce in texture recognition the use of statistical parametric models of the dependence between descriptors. In this work, we chose Hidden Markov Models (HMM) which are both well statistically-based and appropriate models for such a task. They are parametric models and their use requires non trivial parameter estimation. We propose to use recent estimation procedures based on the mean field principle of statistical physics. Using sample images, textures are then learned as HMM's and a set of estimated parameters is associated to each texture. At recognition time, another HMM is used to compute, for each feature vector, the membership probabilities to the different texture classes. Preliminary experiments show promising results [22] .

6.3. Semi and non parametric methods

6.3.1. *Modelling extremal events*

Participants: Stéphane Girard, Laurent Gardes.

Joint work with Mhamed El Aroui (ISG, Tunis), Myriam Garrido (ENAC, Université Toulouse 3, Jean Diebolt (CNRS)).
The first part of our work is to propose new estimates of the extremal index. This parameter is important in practice since it drives the behaviour of the distribution tail. The second part is then to deduce estimates for extreme quantiles.

In [],[12], we investigate the asymptotical behaviour of two new estimates based on double threshold methods.

We also introduce a quasi-conjugate Bayes approach for estimating Generalized Pareto Distribution (GPD) parameters, distribution tails and extreme quantiles within the Peaks-Over-Threshold framework [24]. Bayes credibility intervals are defined, they provide assessment of the quality of the extreme events estimates. Posterior estimates are computed by Gibbs samplers with Hastings-Metropolis steps. Even if non-informative priors are used in this work, the suggested approach could incorporate informative priors. It brings solutions to the problem of estimating extreme events when data are scarce but expert opinion is available.

Finally, we introduce estimates dedicated to the important case of Weibull tail-distributions [9][13] which includes for instance Gaussian, gamma, and Weibull distributions.

6.3.2. Boundary estimation

Participants: Stéphane Girard, Laurent Gardes.

Joint work with Anatoli Iouditski (Univ. Joseph Fourier, Grenoble), Pierre Jacob, Ludovic Menneteau (Univ. Montpellier) and Alexandre Nazin (IPU, Moscow, Russia).
The first part of our work consists in building nonparametric estimates of the boundary of some support based on the extreme values of the sample [30][29][27], [17][16]. These estimates require to select which extreme values are to be used. This problem is difficult in practice. To overcome this limitation, estimates based on a linear programming formulation are defined. In this case, the important points of the sample are selected automatically by solving a linear optimization problem [1][11]. Our current work consists in building an optimization problem leading to an optimal estimate for the L_1-distance. We refer to [15] and [14] for similar studies on other estimates.

6.3.3. Sensitivity analysis and model uncertainty

Participants: Julien Jacques, Christian Lavergne.

Joint work with Nicolas Devictor (CEA - Cadarache).
The first motivation of J. Jacques thesis was to take into account model uncertainty in sensitivity analysis. Two types of uncertainty have been studied: uncertainty due to the use of a simplified model and uncertainty du to a mutation of the model. A second motivation was exhibited during the first thesis year: the problem of sensitivity analysis of models with correlated inputs.
This last year of thesis has been devoted to the formalisation of the proposed solutions and to several applications in nuclear engineering.
This thesis work has been presented at the fourth international conference on Sensitivity Analysis of Model Output, and at two others French conferences. A paper has been accepted in the journal *Reliability Engineering and System Safety* [18].

6.3.4. Dimension reduction for image processing

Participants: Stéphane Girard, Charles Bouveyron.

Joint work with Serge Iovleff (Université Lille 3) and Cordelia Schmid (Lear, Inria).
In the first part of this work, we focus on nonlinear PCA based on manifold approximation of the set of points introduced in [8]. This method proves especially useful when the observations are images [4] and thus located in high dimensional spaces. The joint work with Serge Iovleff consists in defining a probabilistic framework for nonlinear PCA permitting new extensions of this dimension-reduction method [28].

The second part of our work is to propose new methods combining dimension-reduction with a classification step[20]. This is the context of the PhD thesis of Charles Bouveyron which takes place in collaboration with C. Schmid (Lear) in the ACI Movistar in the “Masse de données” program. A new method of discriminant analysis, called High Dimensional Discriminant Analysis (HHDA) is introduced. Our approach is based on the assumption that high dimensional data live in different subspaces with low dimensionality. Thus, HDDA reduces the dimension for each class independently and regularizes class conditional covariance matrices in order to adapt the Gaussian framework to high dimensional data. This regularization is achieved by assuming that classes are spherical in their eigenspace.

6.3.5. Sparse Continuous Wavelet Transform Inversion

Participant: Paulo Gonçalves.

Joint work with P. Borgnat (Inria post-doctoral fellowship).

This ongoing work, initiated with P. Borgnat during his post-doctoral stay at IST-ISR (sept. 2003 – sept. 2004), aims at recovering a signal from the sparse set of local maxima coefficients of its wavelet decomposition. Starting with the conjugate gradient algorithm proposed by Mallat and Zhong to pseudo-inverse the transform, we adapted it to complex wavelets. There are two main advantages in using complex wavelets for this purpose:

1. the number of local maxima is considerably reduced when considering the magnitude of the complex wavelet transform field, as compared to its real part.
2. Although the reconstruction error is slightly smaller with real wavelets, in most case, it decreases faster with complex wavelets.

With J. Lewalle (Univ. of Syracuse, New York, USA), we are now tackling the continuous wavelet inversion problem from the point of view of its diffusion formulation (PDE).

6.3.6. Diffusion of time-frequency representations

Participant: Paulo Gonçalvès.

This topic is at the core of J. Gosme (Univ. Tech. Troyes) Ph.D. thesis (to be defended on December 20, 2004) advised by C. Richard (Univ. Tech. Troyes) and co-advised by P. Gonçalvès (INRIA).
Our aim is to propose a totally adaptive (signal driven) smoothing of time-frequency representations, relying on non linear anisotropic diffusion schemes inspired from the heat equation. We derived a set of partial differential equations applied to standard time-frequency representations (e.g. Wigner-Ville distribution) to locally adapt the amount of smoothing to the local (time-frequency) characteristics of the signal. The outcomes are for instance interference free representations with sharp localization properties, but the versatility of this approach allows for enhancing any other desired feature of the distributions, defining a corresponding diffusion control strategy (conductance function). An important achievement this year, was to derive an equivalent diffusion process that preserves covariances with respect to time shifts and scale changes, opening up in this way the scope of adaptive smoothing to the affine class of time-scale representations.

6.3.7. Empirical Mode Decomposition

Participant: Paulo Gonçalvès.

This topic is the main line of our scientific collaboration with Ecole Normale Superieure de Lyon (France). P. Flandrin and P. Goncalvès are co-advising the PhD thesis of G. Rilling (starting date, Sept. 2004) on “Empirical Mode Decomposition" (EMD).
We now briefly describe the EMD technique. This entirely data-driven algorithm introduced by N. E. Huang decomposes iteratively a complex signal (i.e. with several characteristic time scales coexisting) into elementary AM-FM type components (Intrinsic Mode Functions). The rationale of this decomposition is to locally identify in the signal the most rapid oscillations, defined as the waveform interpolating interwoven local maxima and minima. To do so, local maxima points (respectively local minima points) are interpolated with a cubic spline, to yield the upper (resp. lower) envelope. The mean envelope (half sum of upper and lower envelopes) is then subtracted from the initial signal, and the same interpolation scheme is re-iterated on the remainder. The so-called *sifting process* stops when the mean envelope is reasonably zero everywhere, and the resulting signal is designated the first *Intrinsic Mode Function*. The higher order IMFs are iteratively extracted applying the same procedure to the initial signal after the previous IMFs have been removed.

With P. Flandrin (ENS-Lyon, France) and G. Rilling (ENS-Lyon, France), we are pursuing the qualitative study of EMD as an adaptive dyadic filter bank. In the course of this analysis we have also proposed several modifications of this decompositions, that significantly improved its performances (cf. corresponding publications).

With S. Bausson (IST-ISR) and P. de Oliveira (marinha & IST-ISR), we are continuing a work that P. Goncalvès had initiated at Inria with B. Esterni, a post-graduate student from Ensimag (France). We endeavored to transpose the EMD to 2D signals, and more specifically to quadratic time-frequency representations of 1D signals. The idea is to use EMD to separate signal components (low pass structures) from cross-components (high pass oscillating terms).

In parallel to this, we are investigating several different approaches to the 2D-EMD, including for instance a row-wise/column-wise decomposition, in the spirit of the so-called *non-standard wavelet transform*. This is also a joint work with J.C. Nunes (Université de Créteil, France).

6.3.8. Statistical Modelling of Image Symmetries and Stationarization

Participant: Paulo Gonçalvès.

Joint work with P. Borgnat. This research topic was prompted by the tight connection between the work of P. Borgnat developed during his PhD thesis (ENS-Lyon, Nov. 2002) and the current activities on local stationarity of Professors I. Lourtie (IST-ISR) and F. Garcia (IST-ISR). For timetable issues, the achievement of this work has been delayed, but should remain the backbone of a collaboration between INRIA, IST-ISR and Ecole Normale Supérieure de Lyon (France).

The proposed work deals with 2D statistical fields, for instance images but also other random fields coming from other domains (e.g., in physics, the turbulent velocity fields, or a representation of a 1D signal on a time-frequency plane). Knowing how to define the symmetries of one image is a classical way to describe textures (leaving out the study of shapes for now).

Among the interesting symmetries, the scale invariance property has a special relevance both for images (to deal with multi-scale structures) and physical fields. The first part of this work was to define what are the possible choices of symmetries for images, especially in the case of scale invariance (or self-similarity for random fields). Using preliminary work on plane transformations, we have studied how one can use a stationarization of those invariances to prescribe the statistical properties of the random fields. Stationarization is a method that studies a signal or field that has some invariance by means of a stationary generator. Namely, one tries to find a stationary generator $Y(t)$ that can be warped by some warping $t = f(u)$ in the original field $X(u) = Y(f(u))$ that has a different invariance. This method was introduced in geostatistics and used in some problems of imaging. We develop this approach for self-similarity of images.

A first point was to describe possible warping functions and the kinds of self-similarity that can be targeted this way. The correlation structure is then controlled by the invariance. We have studied how using the stationary generator (and thus, means to synthesize this field Y using this stationarity – spectral or parametric methods) induces an efficient method for the synthesis of self-similar random fields. A second point is the question of analysis: is it possible to recover the stationarizing warping from one realization of the random field ? Drawing on the method proposed by Perrin and Senoussi (1999) based on the variogram, and on the work of Clerc and Mallat (2000) on wavelet decompositions, we address the problem of scale invariant fields. Preliminary results show that it is possible in this case to recover the warping but a more robust method should be designed. An insight would be to adapt results about local stationarity (work of F. Garcia and I. Lourtie at the ISR) to cross-check the stationarity of the unwarped process locally, during the estimation of the inverse warping.

This work was presented in a workshop at INRIA Rocquencourt in December 2003 (*journées Thalweg*).

6.4. Reliability

6.4.1. An Aging model

Participant: Henri Bertholon.

Joint work with G. Celeux and N. Bousquet. In the reliability context, we are interested in lifetime data analysis. We have especially examined a simple competing risk model that may be viewed as a possible alternative to the standard Weibull model. In particular our model enables to take into account both accidental causes of failure and aging. The estimation of parameters is made by Maximum Likelihood and Bayesian inference. Moreover in order to discriminate between our model and Weibull (or exponential) models, a test procedure has been proposed. Finally different applications have been presented.

7. Contracts and Grants with Industry

7.1. Sensitivity analysis and model uncertainty

Participants: Julien Jacques, Christian Lavergne.

This contract with the LCFR (Laboratoire de Conduite et Fiabilité des Réacteurs) of CEA/Cadarache/DER concerned sensitivity analysis and model uncertainty. It funded during three years the thesis of Julien Jacques.

8. Other Grants and Activities

8.1. Regional initiatives

MISTIS participates in the weekly statistical seminar of Grenoble, F. Forbes is one of the organizers and several lecturers have been invited in this context.

8.2. National initiatives

MISTIS got a Ministry grant (Action Concertée Incitative Masses de données) for a three-year project involving other partners (Team Lear from INRIA, SMS from University Joseph Fourier and Heudiasyc from UTC, Compiègne). The project called Movistar aims at investigating visual and statistical models for image recognition and description and learning techniques for the management of large image databases.

8.3. International initiatives

8.3.1. Europe

P. Gonçalvès is since September 1st, 2003 on leave at *Instituto de Sistemas e Robotica* of *Instituto Superior Tecnico*, Lisbon (Portugal).
S. Girard is a member of the European project (Interuniversity Attraction Pole network) "Statistical techniques and modelling for complex substantive questions with complex data",
Web site : http://www.stat.ucl.ac.be/IAP/frameiap.html.

S. Girard has also joint work with Prof. A. Nazin (Institute of Control Science, Moscow, Russia).

8.3.2. North Africa

C. Lavergne and F. Forbes are involved in a one-year project (STIC-INRIA-universités tunisiennes) with other INRIA teams and ISG Tunis (Institut Superieur de Gestion). C. Lavergne is supervising M. Saidane as a PhD student.

S. Girard has also joint work with M. El Aroui (ISG Tunis).

8.3.3. North America

F. Forbes has joint work with:
- C. Fraley (Univ. of Washington, USA)
- A. Raftery (Univ. of Washington, USA)

P. Gonçalvès has joint work with:
- R. Riedi (Rice Univ., USA)
- R. Baraniuk (Rice Univ., USA)
- A. Feuerverger (Univ. of Toronto, CA).
- J. Lewalle (Univ. of Syracuse, USA).

8.4. Visiting scientists

Prof. Alexandre Nazin from Institute of Control Science, Moscow, spent two months in the team.

9. Dissemination

9.1. Leadership within scientific community

C. Lavergne is member of the "Institut de Mathématiques et de Modélisation", Montpellier, UMR CNRS 5149.

S. Girard defended his HDR thesis in July 2004 entitled *Contributions à l'inférence statistique semi- et non-paramétrique*.

S. Girard reported on the PhD thesis of Imen Rached from university Marne-La-Vallée, entitled *Moments pondérés généralisés*.

F. Forbes was co-organizer of the 5th French Danish workshop on "Spatial Statistics and image analysis in biology" held in Saint Pierre de Chartreuse (France), from May 10 to 13, 2004.

S. Girard was chairman for the Third International Symposium on Extreme Value Analysis 2004 (Portugal), and for the 36emes Journées de Statistique (Montpellier in May 2004).

P. Gonçalvès was director (and co-organizer) of the "Wavelet And Multifractal Analysis" summer school held in Cargèse (Corsica, France) from July 19 to 31, 2004.

9.2. University Teaching

F. Forbes lectured a graduate course on statistics at Poly Tech, Univ. J. Fourier, Grenoble.

L. Gardes, S. Girard are faculty members at Univ. P. Mendes France and Univ. J. Fourier in Grenoble. C. Lavergne is professor in Montpellier and H. Berthelon is faculty member at CNAM, Paris.

10. Bibliography

Major publications by the team in recent years

[1] G. BOUCHARD, S. GIRARD, A. IOUDITSKI, A. NAZIN. *Nonparametric Frontier estimation by linear programming*, in "Automation and Remote Control", vol. 65, n° 1, 2004, p. 58–64.

[2] G. CELEUX, S. CHRÉTIEN, F. FORBES, A. MKHADRI. *A Component-wise EM Algorithm for Mixtures*, in "Journal of Computational and Graphical Statistics", vol. 10, 2001, p. 699–712.

[3] G. CELEUX, F. FORBES, N. PEYRARD. *EM Procedures Using Mean Field-Like Approximations for Markov Model-Based Image Segmentation*, in "Pattern Recognition", vol. 36, n° 1, 2003, p. 131-144.

[4] B. CHALMOND, S. GIRARD. *Nonlinear modeling of scattered multivariate data and its application to shape change*, in "IEEE Pattern Analysis and Machine Intelligence", vol. 21, n° 5, 1999, p. 422–432.

[5] P. FLANDRIN, P. GONÇALVÈS, P. ABRY. *Lois d'échelle, Fractales et Ondelettes*, Traité Information - Commande - Communication, vol. 2, chap. Analyses en ondelettes et lois d'échelle, Abry, P., Gonçalvès, P. and Lévy Véhel, J. eds, Paris, France, 2002.

[6] F. FORBES, N. PEYRARD. *Hidden Markov Random Field Model Selection Criteria based on Mean Field-like Approximations*, in "in IEEE trans. PAMI", August 2003.

[7] F. FORBES, A. E. RAFTERY. *Bayesian Morphology: Fast Unsupervised Bayesian Image analysis*, in "Journal of the American Statistical Association", vol. 94, nº 446, June 1999, p. 555-568.

[8] S. GIRARD. *A nonlinear PCA based on manifold approximation*, in "Computational Statistics", vol. 15(2), 2000, p. 145-167.

[9] S. GIRARD. *A Hill type estimate of the Weibull tail-coefficient*, in "Communication in Statistics - Theory and Methods", vol. 33, nº 2, 2004, p. 205–234.

[10] C. ROBERT. *Méthodes statistiques pour l'I.A. ; l'exemple du diagnostic médical*, Masson, Paris, 1991.

Articles in referred journals and book chapters

[11] G. BOUCHARD, S. GIRARD, A. IOUDITSKI, A. NAZIN. *Linear programming problems for frontier estimation*, in "Applied Stochastic Models in Business and Industry", To appear, 2004.

[12] L. GARDES, S. GIRARD. *Asymptotic properties of a Pickands type estimator of the extreme value index*, in "Focus on probability theory, New-York", F. COLOMBUS (editor)., to appear, Nova Science, 2004.

[13] L. GARDES, S. GIRARD. *Estimating extreme quantiles of Weibull tail-distributions*, in "Communication in Statistics - Theory and Methods", to appear, 2004.

[14] S. GIRARD. *On the asymptotic normality of the L_1- error for Haar series estimates of Poisson point processes boundaries*, in "Statistics and Probability Letters", vol. 66, 2004, p. 81–90.

[15] S. GIRARD, P. JACOB. *Asymptotic normality of the L_1-error for Geffroy's estimate of Poisson point process boundaries*, in "Publications de l'Institut de Statistique de l'Université de Paris", to appear, 2004.

[16] S. GIRARD, P. JACOB. *Extreme values and kernel estimates of point processes boundaries*, in "ESAIM: Probability and Statistics", vol. 8, 2004, p. 150–168.

[17] S. GIRARD, L. MENNETEAU. *Central limit theorems for smoothed extreme value estimates of point processes boundaries*, in "Journal of Statistical Planning and Inference", to appear, 2004.

[18] J. JACQUES, C. LAVERGNE, N. DEVICTOR. *Sensitivity Analysis in presence of model uncertainty and correlated inputs*, in "Reliability Engineering and System Safety", To appear, 2004.

Internal Reports

[19] F. FORBES, N. PEYRARD, C. FRALEY, D. GEORGIAN-SMITH, D. GOLDHABER, A. E. RAFTERY. *Model-based Region of Interest Selection in Dynamic Breast MRI*, Technical report, nº 472, Stat. dept, Univ. of Washington, 2004.

Miscellaneous

[20] C. BOUVEYRON, S. GIRARD, C. SCHMID. *High Dimensional Discriminant Analysis*, submitted for publication, 2004.

[21] J. DIEBOLT, M. GARRIDO, S. GIRARD. *A goodness-of-fit test for the distribution tail*, submitted for publication, 2004.

[22] F. FORBES, J. BLANCHET. *Markov Random Fields for Recognizing Textures modeled by Feature Vectors*, submitted for publication, 2004.

Bibliography in notes

[23] J. DIEBOLT, J. ECARNOT, M. GARRIDO, S. GIRARD, D. LAGRANGE. *Le logiciel Extremes, un outil pour l'étude des queues de distribution*, in "La revue de Modulad", vol. 30, 2003, p. 53–60.

[24] J. DIEBOLT, M. EL-AROUI, M. GARRIDO, S. GIRARD. *Quasi-conjugate Bayes estimates for GPD parameters and application to heavy tails modelling*, Technical report, n° RR-4803, INRIA, 2003, http://www.inria.fr/rrrt/rr-4803.html.

[25] F. FORBES, N. PEYRARD, C. FRALEY, D. GEORGIAN-SMITH, D. GOLDHABER, A. E. RAFTERY. *Region of interest selection and dynamic breast MRI data Analysis using multivariate statistical methods for Clustering and spatial segmentation*, Technical report, n° RR-4249, Inria Rhone-Alpes, 2001, http://www.inria.fr/rrrt/rr-4249.html.

[26] G. FORT, E. MOULINES. *Convergence of the Monte-Carlo EM for curved exponential families*, in "Annals of Statistics", vol. 31, n° 4, 2003, p. 1220–1259.

[27] L. GARDES. *Estimating the support of a Poisson process via the Faber-Shauder basis and extreme values*, in "Publications de l'Institut de Statistique de l'Université de Paris", vol. XXXXVI, 2002, p. 43-72.

[28] S. GIRARD, S. IOVLEFF. *Auto-Associative Models and Generalized Principal Component Analysis*, in "Journal of Multivariate Analysis", vol. 93, n° 1, 2005, p. 21–39.

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

[30] S. GIRARD, P. JACOB. *Projection estimates of point processes boundaries*, in "Journal of Statistical Planning and Inference", vol. 116, n° 1, 2003, p. 1–15.

[31] J. ZHANG. *The Mean Field Theory in EM Procedures for Markov Random Fields*, in "IEEE Trans. on signal processing", vol. 40, n° 10, 1992, p. 2570–2583.