INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team IS2

Inférence statistique pour l'industrie et la santé

Rhône-Alpes

THEME 4A

Table of contents

1. Team

Head of project-team
Gilles Celeux [DR Inria]

Staff member
Florence Forbes [CR Inria]
Paulo Gonçalvès [CR Inria]
Christian Lavergne [professeur des universités, Montpellier]

Inria Post-doctoral fellow
Edwige Allain
Emilie Lebarbier [until October 2003]

Project technical staff
Jérome Ecarnot [until March 2003]
Grégory Noulin

Ph. D. student
Guillaume Bouchard [Inria]
Charles Bouveyron [MENRT since October 2003]
Julien Jacques [Inria]
Carine Vera [Inra/Cirad, Montpellier]
Matthieu Vignes [AC]
Mohamed Saidane [Teaching Assistant, Tunis]

Research scientist (partner)
Henri Bertholon [Faculty member CNAM, Paris]
Franck Corset [Teaching Assistant, Grenoble]
Jean Diebolt [DR CNRS, Marne-la-Vallée]
Jean-Baptiste Durand [Post-doctoral fellow, Cirad, Montpellier since March 2003]
Myriam Garrido [Teaching Assistant, Grenoble]
Stéphane Girard [maître de conférences, Grenoble]
Olivier Martin [CR Inra, Montpellier since February 2003]

Administrative assistant
Françoise de Coninck
Élodie Toihein [since October 2003]

2. Overall Objectives

The team IS2 aims at doing research in statistical modelling for industrial and medical fields. It deals with diagnosis, reliability, industrial failure time analysis, models for hospital lengths of stay, statistical image analysis and finance. The chosen methods involve generalized linear models, hidden structure models identified by stochastic algorithms, adaptative estimation, conditionally heteroscedastic models for time series, dynamical systems identification.

3. Scientific Foundations

3.1. Hidden structure models

Participants: Guillaume Bouchard, Gilles Celeux, Jean-Baptiste Durand, Florence Forbes, Paulo Gonçalvès, Christian Lavergne, Emilie Lebarbier, Olivier Martin, Matthieu Vignes.

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

Hidden structure models are useful for taking into account heterogeneity in data. They concern many domains of statistical methodology (finite mixture analysis, hidden Markov models, random effect models, ...). Owing to their missing data structure, they involve specific difficulties for both estimating the model parameters and assessing its performance. The team IS2 is concerned with 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.

3.2. Random effects and heteroscedasticity

Participant: Christian Lavergne.

Key words: *Generalized linear models*, *heteroscedasticity*, *random effects models*, *ARCH models*.

Regression aims at modelling the relation between a response variable and some regressors. Multiple regression models propose generalizations of the classical linear model like:

- Mixed linear models
- Generalized linear models
- ARCH models and GLM-ARCH
- Generalized linear mixed models.

3.3. Wavelets and Scaling laws

Participant: Paulo Gonçalvès.

Key words: *estimation*, *scaling laws*, *wavelets*, *singularity spectra*.

Our goal here 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.*mon-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. Reliability

Participants: Henri Bertholon, Gilles Celeux, Franck Corset, Julien Jacques, Christian Lavergne, Myriam Garrido, Stéphane Girard.

An important domain of application for IS2 concerns Reliability and industrial lifetime analysis. This activity is developed essentially through collaborations with the EDF research department and the laboratory LCFR of CEA / Cadarache.

4.2. Statistique biomédicale

Participants: Gilles Celeux, Paulo Gonçalvès, Olivier Martin, Christian Lavergne.

A secondary domain of applications concerns biomedical statistics and molecular biology.

5. Software

5.1. MixMod (Mixture Modelling) freeware

Participants: Gilles Celeux, 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.htm`.

5.2. The Extremes freeware

Participants: Jean Diebolt, Jérôme Ecarnot, Myriam Garrido, Stéphane Girard.

The EXTREMES software is a toolbox dedicated to the modelling of extremal events offering extreme quantile estimation procedures and model selection methods [18]. This software is due to 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.

6. New Results

6.1. Hidden structure models

6.1.1. Importance sampling for hidden structure models

Participant: Gilles Celeux.

Joint work with Jean-Michel Marin and Christian Robert (Ceremade, Paris Dauphine).
Missing variable models are typical benchmarks for new computational techniques in that the ill-posed nature of missing variable models offers a characteristic testing ground for these techniques. Their special features also allow for an easier calibration of most computational techniques, by virtue of the data completion they naturally offer. The potential of this approach and its specifics in missing data problems has been studied in settings of increasing difficulty, in comparison with existing approaches.

6.1.2. Hidden Markov models for objects tracking

Participant: Gilles Celeux.

Joint work with Jorge Marques and Jacinto Nascimento (ISR-IST, Lisbonne).
Object tracking problems can be faced by trying to describe the motion of the object with a set of significant dynamical models capable to describe it in a reliable way. We propose an hidden Markov model for object tracking. An exact technique to estimate the models involved in the objects motion has been conceived. Learning is achieved via an EM algorithm. We extend the study to determine a reliable number of models when the motion contains model uncertainty with BIC criterion.

6.1.3. Mixed multiphasic models for longitudinal data analysis

Participants: Carine Véra, Christian Lavergne.

Joint work with Yann Guédon (Cirad - Montpellier).
Preliminary studies show that the statistical modelling of the whole tree growth period with a single linear mixed model was not satisfactory: the model was difficult to interpret because of its high number of parameters. Moreover, the exploratory analysis of tree growth data shows that tree growth period is made up of a succession of growth phases. Consequently, we are developing a new family of statistical models : mixed multiphasic models. They result from the combination of two types of models:

- a Markov chain which models the succession of growth phases,
- linear mixed models associated with each state of the underlying Markov chain, each linear mixed model modelling a growth phase, with climatic covariables and individual random effects.

6.2. Model selection criteria

6.2.1. Penalized likelihood criteria

Participants: Gilles Celeux, Florence Forbes, Emilie Lebarbier.

Joint work with Frédérique Letué (Grenoble).
We organized a working group to compare classical model selection criteria as BIC with Random criteria proposed by Birgé and Massart from a non asymptotic point of view. E. Labarbier has studied, with this approach, the problem of detecting the change-points in mean of a signal corrupted by an additive Gaussian noise. Moreover, a procedure for selecting the number of the clusters by extending the results of E. Lebarbier, in a non supervised context is in progress.

6.2.2. DIC criteria for hidden structure models

Participants: Gilles Celeux, Florence Forbes.

Joint work with Mike Titterington (Glasgow, Scotland) and Christian Robert, (CEREMADE, Paris Dauphine). The deviance information criterion (DIC) introduced by Spiegelhater et al. is directly inspired by linear and generalized linear models, but it is not so naturally defined for missing data models. We have reassessed the criterion for such models, testing the behavior of various extensions in the cases of mixture and random effect models.

6.2.3. Spherical Gaussian mixtures for supervised classification

Participants: Guillaume Bouchard, Gilles Celeux.

Linear Discriminant Analysis is a reference method in supervised classification. In cases where it performs poorly, alternative methods are required. We propose a method based on estimating the density of each group using a mixture of spherical Gaussian distributions. The features of this model are flexibility, simplicity and parsimony. Moreover, we propose choosing the numbers of mixture components with a Bayesian Entropy Criterion which is a penalized likelihood criterion taking into account the classification task.

6.3. Reliability

6.3.1. An Ageing model

Participants: Henri Bertholon, Gilles Celeux.

With the Ensimag student, Nicolas Bousquet, we studied this year Bayesian Inference for the Bertholon distribution, a competing risk model involving an Exponential and a Weibull distributions. This Bayesian inference has been performed through importance sampling schemes taking profit of the missing structure data of the model.

6.3.2. Inverse problems in Reliability

Participants: Edwige Allain, Gilles Celeux.

We studied simulated scenarii to detect the reasons of flaws or incidents on nuclear plants via statistical inference. We essentially made use of mixture analysis and statistical tests. We used expert opinions to decide between alternative flaw scenarii.

6.3.3. Modelling extremal events

Participants: Jean Diebolt, Myriam Garrido, Stéphane Girard.

Joint work with Mhamed El Aroui (ISG, Tunis).
We introduce a quasi-conjugate Bayes approach for estimating Generalized Pareto Distribution (GPD) parameters, distribution tails and extreme quantiles within the Peaks-Over-Threshold framework [40]. 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.

6.3.4. Sensitivity analysis and model uncertainty

Participants: Julien Jacques, Christian Lavergne.

Joint work with Nicolas Devictor (CEA - Cadarache).
During this second year of J. Jacques thesis, two research orientations, which were exhibit during the first year, have been explored. The first concern the impact of model uncertainty on results of sensitivity analysis. This uncertainty, which can be due to the use of a simplified model, or to a process mutation, is considered like a mutation of the start model.we have therefore listed all possible mutations, and analyzed the impact on sensitivity indices. Applications are in progress. The second research orientation concerns sensitivity analysis for models with non independent inputs. Investigations have led to introduce multidimensional sensitivity indices, which are under study.

6.4. Biometric

Participants: Gilles Celeux, Florence Forbes, Paulo Gonçalvès, Olivier Martin, Christian Lavergne, Matthieu Vignes.

6.4.1. Stastistical tools for the analysis of bacterial genomes organisation

Participants: Florence Forbes, Matthieu Vignes.

Joint work with Alain Viari (Inria Rhône-Alpes) and Eduardo Rocha, (Institut Pasteur-ABI, Paris).
We investigated a part of the exploratory analysis of bacterial genomes, beyond gene detection. Our goal is to link proximities among genes on the chromosome with genetic mechanisms of the cell. We reviewed the main work in progress on the subject in order to suggest a formalism best possible suited. We focused on the notion of neighborhood in a broad sense which leads to some specific mathematical tools like renewal process. We certainly need to interact with biologists who are the only in a position to judge if our work is judicious.

6.4.2. Mixture of linear mixed models for clustering gene expression profiles

Participants: Gilles Celeux, Christian Lavergne, Olivier Martin.

Data variability can be important in micro-array data analysis. Thus, when clustering gene expression profiles, it could be judicious to make use of repeated data. In this work, the problem of analyzing repeated data in the model-based cluster analysis context is considered. Linear mixed models are chosen to take into account data variability. A mixture of these models are considered. This leads to a large range of possible models depending on the assumptions made on both the covariance structure of the observations and the mixture model. The maximum likelihood estimation of this family of models through the EM algorithm is presented. The problem of selecting a particular mixture of linear mixed models is considered using penalized likelihood criteria. Illustrative Monte Carlo experiments are presented and an application to the clustering of gene expression profiles is detailed. All those experiments highlight the interest of linear mixed model mixtures to take into

account data variability in a cluster analysis context.They also show encouraging results of the BIC criterion for selecting a relevant model.

6.4.3. Double-blind multicentric study in Parkinson's disease

Participant: Christian Lavergne.

Joint work with Pierre Pollak (Department of Neurology, Grenoble University Hospital).
Detailed title : Double-blind multicentric study of bilateral subthalamic nucleus deep brain stimulation in Parkinson's disease.
After the introduction of levodopa therapy, many patients with idiopathic Parkinson's disease (PD) develop progressive disabling motor complications whose clinical, social and economic impacts impair health-related quality of life (HR-QOL). Deep brain stimulation of the subthalamic nucleus (STN) is an alternative surgical therapy to medical treatment but its initial high cost has limited its diffusion for many years. The French SPARK study group conducted a prospective multicentric study of STN stimulation in advanced PD in order to assess safety, efficiency but also social and economic impacts of this technique.

6.4.4. Statistical modelling of cardiovascular data

Participant: Paulo Gonçalvès, Christian Lavergne.

Joint work with Christophe Lenoir and Bernard Swynghedauw (Inserm, Paris).
This study deals with the heart beat rate analysis of mammalians. In particular, we investigate the role of the autonomous nerve system in mice hearts. In this direction, we have been led to identify and characterize the action of pharmacological autonomic blockades (propranolol and atropine) on the baseline heart rate. To cope with the nature of the experimental setup (repeated and incomplete measures), we then resorted to statistical mixed effect models. An article presenting the results of our study has been submitted to Cardiovascular Research.

6.5. Statistical inference for signal and image processing

6.5.1. Principal Component Analysis and image processing

Participant: Stéphane Girard.

Joint work with Serge Iovleff (Lille).
We focus on nonlinear PCA based on manifold approximation of the set of points introduced in [9]. This method reveals especially useful when the observations are images 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 method.

6.5.2. Boundary estimation

Participants: Guillaume Bouchard, Stéphane Girard.

Joint work with Anatoli Iouditski (Imag, Grenoble) ; Pierre Jacob, Ludovic Menneteau (Montpellier) and Alexandre Nazin (IPU, Moscou, Russie).
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 [26][25][27]. 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 [14].

6.5.3. Empirical Mode Decomposition

Participant: Paulo Gonçalvès.

Joint work with P. Flandrin (CNRS, ENS Lyon) and P. Oliveira (IST-ISR, Lisbonne).
Empirical Mode Decomposition (EMD) is a complex signal analysis algorithm, recently proposed by Huang et al. (1998). Because so far no theory seems appropriate to mathematically formalize the concept of EMD, only

extensive numerical simulations have been performed in order to assess the method [22][35][34]. In parallel, with Benjamin Esterni (a graduate student inter), we developed a bi-dimensional extension of EMD for image processing purposes [43].

6.5.4. Existence test of moments

Participant: Paulo Gonçalvès.

Joint work with Rudolf Riedi (Rice university, Houston (TX), USA).
In this study, we theoretically stated and demonstrated the relation existing between the decay of a fat tail probability distribution of a random variable, and the finiteness of its high order moments. We also proposed an efficient and simple wavelet based estimator for determining the bounds of existing moments, given a finite sample size of unknown random variables.

6.5.5. Wavelet based Stable law regression

Participant: Paulo Gonçalvès.

Joint work with Anestis Andoniadis (Imag - Grenoble) and Andrey Feuerverger (University of Toronto, Canada).
In this work, we exploited the sparsity of wavelet representations to propose an α-stable law regression. Given N i.i.d. realizations of a stable variable, we propose to fit the empirical characteristic function (ECF) through its pruned wavelet decomposition. Indeed, as most non-zero wavelet coefficients are lying beneath the singularity of the ECF at the origin, we can restrict the fitting procedure to these very few coefficients only. We end up with a sparse matrix which is much easier to invert, and a speeded up stable parameters estimation procedure.

6.5.6. Conditional heteroscedasticity dynamic factor models

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.

7. Contracts and Grants with Industry

7.1. Flaw detection in PWR vessel

Participants: Guillaume Bouchard, Gilles Celeux.

This "CRECO" contract with the Reliability group of EDF R&D Chatou concerned Bayesian modelling and inference through MCMC methods of the statistical distributions of flaws PWR vessels.

7.2. Likelihood in flaw causal scenarii

Participants: Edwige Allain, Gilles Celeux.

This "CRECO" contract with the Reliability group of EDF R&D Chatou concerned statistical inference regarding the study of material features provoking flaws in nuclear equipments.

7.3. Extremal events

Participants: Jean Diebolt, Jérôme Ecarnot, Myriam Garrido, Stéphane Girard.

This contract with the Reliability group of EDF R&D Chatou concerned statistical inference for extremal events. It funded the development of the EXTREMES software.

7.4. 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

IS2 participates in the weekly statistical seminar of Grenoble, G. Celeux is one of the organizers and several lecturers have been invited in this context.
P. Gonçalvès is with two thematic regional programs: « Application de l'Analyse en Ondelettes à l'Acoustique et à la Turbulence » headed by V. Perrier (ENSIMAG-INPG) and « Diagnostic Acoustique de la Vorticité dans les Écoulements Turbulents » headed by C. Baudet (LEGI-UJF). Both programs have reached their third and last year of existence.
P. Gonçalvès is a member of the research IMAG group, entitled « Analyse Multirésolution, Ondelettes et Applications », headed by V. Perrier (ENSIMAG-INPG).
The activities of the FIMA group of reliability continued.It is a collaboration between LMC-SMS and IS2. (http://www-lmc.imag.fr/lmc-sms/Olivier.Gaudoin/FIMA/accueil.html)

8.2. National initiatives

G. Celeux left Inria Rhône-Alpes to join Inria Futurs on September 1st. He is aiming to create in Orsay a new Inria team in association with the Statistical and Probability team of the Mathematical Department of the University Paris-Sud. This Inria Team, called SELECT, will be essentially involved with model selection in statistical learning.
C. Lavergne is member of the "Institut de Mathématiques et de Modélisation", Montpellier, UMR CNRS 5149. He supervised a PhD on the generalized linear mixed model and choice of criteria selection.
P. Gonçalvès has a collaboration with the U572 team of INSERM at the Lariboisière hospital (Paris). In particular, he co-supervised a PhD work on the mice heart rate analysis.

8.3. International initiatives

8.3.1. Europe

P. Gonçalvès is since September 1st on leave at *Instituto de Sistemas e Robótica* of *Instituto Superior Tecnico*, Lisbon (Portugal).
G. Celeux has a collaboration with this institution and he was referee of the PhD Thesis of Jacinto Nascimento in January 2003.
G. Celeux has joint work with M. Tiiterington (Glasgow University).

8.3.2. North America

P. Gonçalves has joint work with:
- Riedi (Rice Univ., USA)
- A. Feuerverger (Univ. of Toronto, CA).

9. Dissemination

9.1. Leadership within scientific community

P. Gonçalvès is an Associate Editor of *IEEE Signal Processing Letters*.

P. Gonçalvès is co-organizing the "Wavelet And Multifractal Analysis" summer school to be held in Cargèse (Corsica, France) from July 19th to 31st, 2004.
G. Celeux was invited lecturer for the INSERM workshop on “Statistical methods for microarray data” (Lalonde, May 2003).

9.2. University Teaching

G. Celeux lectured multidimensional statistics in the DEA MIMB, UJF university of Grenoble.
P. Gonçalvès lectured a graduate course on *Time-Frequency and Multi-resolution Analysis* at ENSERG.

9.3. Conference and workshop committees, invited conferences

G. Celeux was invited speaker at the CLADAG meeting in Bologna (September 2003).

10. Bibliography

Major publications by the team in recent years

[1] M. BACHA, G. CELEUX, E. IDÉE, A. LANNOY, D. VASSEUR. *Estimation de modèles de durées de vie fortement censurés.* Eyrolles, Paris, 1998.

[2] C. BIERNACKI, G. CELEUX, G. GOVAERT. *Assessing a mixture model for clustering with the integrated completed likelihood.* in « IEEE Trans. on PAMI », 2000, pages 267-272.

[3] G. CELEUX, J. DIEBOLT. *A stochastic approximation type EM algorithm for the mixture problem.* in « Stochastic and Stochastics Reports », volume 41, 1992, pages 119-134.

[4] G. CELEUX, F. FORBES, N. PEYRARD. *EM Procedures Using Mean Field-Like Approximations for Markov Model-Based Image Segmentation.* in « Pattern Recognition », number 1, volume 36, 2003, pages 131-144.

[5] G. CELEUX, G. GOVAERT. *Gaussian parsimonious clustering models.* in « Pattern Recognition », volume 28, 1995, pages 781-793.

[6] P. FLANDRIN, P. GONÇALVÈS, P. ABRY. *Lois d’échelle, Fractales et Ondelettes.* edition Hermès Sciences Publications, series Traité Information - Commande - Communication, volume 2, Abry, P., Gonçalvès, P. and Lévy Véhel, J. eds, Paris, France, 2002, chapter Analyses en ondelettes et lois d’échelle.

[7] F. FORBES, A. E. RAFTERY. *Bayesian Morphology: Fast Unsupervised Bayesian Image analysis.* in « Journal of the American Statistical Association », number 446, volume 94, June 1999, pages 555-568.

[8] O. FRANÇOIS, C. LAVERGNE. *Design for Evolutionary Algorithms - A Statistical Perspective.* in « IEEE Transactions on Evolutionary Computation », volume 5, 2001, pages 129-148.

[9] S. GIRARD. *A nonlinear PCA based on manifold approximation.* in « Computational Statistics », volume 15(2), 2000, pages 145-167.

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

Doctoral dissertations and "Habilitation" theses

[11] F. CORSET. *Aide à l'optimisation de maintenance à partir de réseaux bayésiens et fiabilité dans un contexte doublement censuré.* Ph. D. Thesis, Université Joseph Fourier, Grenoble, 2003, 22 january.

[12] J.-B. DURAND. *Modèles à structure cachée : inférence, sélection de modèles et applications.* Ph. D. Thesis, Université Joseph Fourier, Grenoble, 2003, 31 january.

Articles in referred journals and book chapters

[13] C. BIERNACKI, G. CELEUX, G. GOVAERT. *Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models.* in « Computational Statistics and Data Analysis », volume 41, 2003, pages 561–575.

[14] G. BOUCHARD, S. GIRARD, A. IOUDITSKI, A. NAZIN. *Nonparametric estimation of the support boundary through linear programming.* in « Automation and Remote Control », 2003, To appear.

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

[16] G. CELEUX. *Analyse discriminante.* G. GOVAERT, editor, in « L'analyse des données », series Information-Commande-Communication, Hermes Science, 2003.

[17] G. CELEUX, F. FORBES, N. PEYRARD. *EM Procedures Using Mean Field-Like Approximations for Markov Model-Based Image Segmentation.* in « Pattern Recognition », volume 36, 2003, pages 131-144.

[18] 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 », 2003.

[19] J. DIEBOLT, M. GARRIDO, S. GIRARD. *Asymptotic normality of the ET method for extreme quantile estimation. Application to the ET test.* in « Comptes-Rendus de l'Académie des Sciences, Série I », volume 337, 2003, pages 213–218.

[20] J.-B. DURAND, L. BOZZI, G. CELEUX, C. DERQUENNE. *Analyse de courbes de consommation electrique par chaines de Markov cachées.* in « Revue de Statistique Appliquée », 2004, To appear.

[21] J.-B. DURAND, P. GONÇALVÈS, Y. GUÉDON. *Statistical Inference for Hidden Markov Tree Models and Application to Wavelet Trees.* in « IEEE Transactions on Signal Processing », 2003, To appear.

[22] P. FLANDRIN, G. RILLING, P. GONÇALVÈS. *Empirical Mode Decomposition as a Filter Bank.* in « IEEE, Signal Processing Letters », 2003, To appear in the Feb. 2004 issue.

[23] F. FORBES, N. PEYRARD. *Hidden Markov Random Field Model Selection Criteria based on Mean Field-like Approximations.* in « in IEEE trans. PAMI », 2003, To appear.

[24] S. GIRARD. *A Hill type estimate of the Weibull tail-coefficient.* in « Communication in Statistics - Theory and Methods », volume 33, 2003.

[25] S. GIRARD, P. JACOB. *Extreme values and Haar series estimates of point process boundaries.* in « Scandinavian Journal of Statistics », volume 30, 2003, pages 369–384.

[26] S. GIRARD, P. JACOB. *Projection estimates of point processes boundaries.* in « Journal of Statistical Planning and Inference », volume 116, 2003, pages 1–15.

[27] S. GIRARD, L. MENNETEAU. *Central limit theorems for smoothed extreme value estimates of point processes boundaries.* in « Journal of Statistical Planning and Inference », 2003, To appear.

[28] J. OVARLEZ, P. GONÇALVÈS, R. BARANIUK. *Analyse temps-fréquence quadratique III : La classe affine et autres classes covariantes.* F. HLAWATSCH, F. AUGER, J. OVARLEZ, editors, in « Temps-Frequence : Concepts et outils », Hermès, 2003, To appear.

Publications in Conferences and Workshops

[29] C. BIERNACKI, G. CELEUX, G. GOVAERT, F. LANGNOGNET, G. NOULIN. *Le logiciel Mixmod.* in « 35èmes Journées de Statistique organisées par la Société Française de Statistique », pages 415–418, Lyon, juin, 2003.

[30] G. BOUCHARD, G. CELEUX. *Supervised classification with spherical Gaussian mixtures.* in « CLADAG », pages 75-79, Bologna, septembre, 2003.

[31] G. BOUCHARD, S. GIRARD, A. IOUDITSKI, A. NAZIN. *Linear programming problems for frontier estimation.* in « Second International Control Conference », Moscou, juin, 2003.

[32] G. CELEUX. *Classification de données répétées issues de puces à ADN. Application à l'analyse de profils d'expression.* in « Atelier de formation : 145 analyse statistique des biopuces », 2003.

[33] J. DIEBOLT, J. ECARNOT, M. GARRIDO, S. GIRARD, D. LAGRANGE. *Le logiciel Extremes.* in « 35èmes Journées de Statistique organisées par la Société Française de Statistique », pages 419–422, Lyon, juin, 2003.

[34] P. FLANDRIN, P. GONÇALVÈS. *Sur la décomposition modale empirique.* in « 19ème Colloque GRETSI », Paris (France), September, 2003.

[35] G. RILLING, P. FLANDRIN, P. GONÇALVÈS. *On Empirical Mode Decomposition and its Algorithms.* in « IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing », Grado-Trieste (Italy), June, 2003.

Internal Reports

[36] E. ALLAIN, G. CELEUX. *Stratégies de mise à l'épreuve de modèles physiques par des données de retour d'expérience.* Rapport de fin de contrat Inria-EDF, Décembre, 2003.

[37] G. CELEUX, F. FORBES, C. ROBERT, M. TITTERINGTON. *Deviance Information Criteria for Missing Data Models.* Technical report, number RR-4859, INRIA, Juin, 2003, http://www.inria.fr/rrrt/rr-4859.html.

[38] G. CELEUX, F. FORBES, C. ROBERT, M. TITTERINGTON. *Learning Switching Dynamic Models for Objects Tracking.* Technical report, number RR-4863, INRIA, Juin, 2003, http://www.inria.fr/rrrt/rr-4863.html.

[39] G. CELEUX, J.-M. MARIN, C. ROBERT. *Iterated Importance sampling in Missing Data Problems.* Technical report, number Technical Report, Universités Paris 9-Dauphine, Juin, 2003.

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

[41] J.-B. DURAND, L. BOZZI, G. CELEUX, C. DERQUENNE. *Analyse de courbes de consommation electrique par chaines de Markov cachées.* Technical report, number RR-4858, INRIA, Juin, 2003, http://www.inria.fr/rrrt/rr-4858.html.

Miscellaneous

[42] N. BOUSQUET. *On an ageing model.* 2003.

[43] B. ESTERNI. *Empirical Mode Decomposition.* 2003.

[44] A. IMOUSSATEN. *Un test d'adéquation aux valeurs extrêmes.* 2003.