IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2011

Project-Team CQFD

Quality control and dynamic reliability

IN COLLABORATION WITH: Institut de Mathématiques de Bordeaux (IMB)

RESEARCH CENTER
Bordeaux - Sud-Ouest

THEME
Stochastic Methods and Models

Table of contents

Project-Team CQFD

Keywords: Stochastic Modeling, Statistical Methods, Control Theory, Optimization, Data Analysis, Reliability

1. Members

Faculty Members

François Dufour [Institut Polytechnique de Bordeaux, Team Leader, Professor (Pr), HdR]
Marie Chavent [University Bordeaux Segalen, Assistant Professor (MdC), HdR]
Anne Gégout-Petit [University Bordeaux Segalen, Assistant Professor (MdC)]
Benoîte de Saporta [University Montesquieu Bordeaux IV, Assistant Professor (MdC)]
Jérôme Saracco [Institut Polytechnique de Bordeaux, Professor (Pr), HdR]
Huilong Zhang [University Bordeaux 1, Assistant Professor (MdC)]
Thi Mong Ngoc Nguyen [University Montesquieu Bordeaux IV, ATER, until Sep. 2011]

PhD Students

Romain Azaïs [INRIA, ANR FAUTOCOES grant, since Sept. 2010]
Camille Baysse [Thales Optronique, CIFRE grant, since Nov. 2010]
Adrien Brandejsky [INRIA CORDI-C EADS grant, since Sept. 2009]
Raphaël Coudret [University Bordeaux 1, MENRT grant, since Oct 2010]
Laurent Vézard [CNRS, Region grant, since Oct 2010]

Administrative Assistants

Marie Sanchez [until Aug. 2011]
Nicolas Jahier [since Sep. 2011]

2. Overall Objectives

2.1. Presentation

The core component of our scientific agenda focuses on the development of statistical and probabilistic methods for the modeling and the optimization of complex systems. These systems require mathematical representations which are in essence dynamic and stochastic with discrete and/or continuous variables. This increasing complexity poses genuine scientific challenges that can be addressed through complementary approaches and methodologies:

- Modeling: design and analysis of realistic and tractable models for such complex real-life systems and various probabilistic phenomena;
- Estimation: developing theoretical and computational procedures in order to estimate and evaluate the parameters and the performance of the system;
- Optimization: developing theoretical and numerical control tools to optimize the performance and/or to maintain the system function in operating state.

2.2. Highlights

F. Dufour is scientific advisor for the C.E.A (French council for the atomic energy) in 2011.

F. Dufour has been invited to present a paper at the SIAM Conference on Control and its Applications, Baltimore, USA (July 2011).

F. Dufour has been invited to present a paper at the conference on Markov & semi-Markov Processes & Related Fields conference, Greece (September, 2011).

The ANR project ADAPTEAU has been obtained for the period 2012-2016 and will start in january 2012.

The team CQFD will organize the first french-speaking meeting on the software R in July 2012.

3. Scientific Foundations

3.1. Introduction

The scientific objectives of the team are to provide mathematical tools for modeling and optimization of complex systems. These systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses genuine scientific challenges in the domain of modeling and optimization. More precisely, our research activities are focused on stochastic optimization and (parametric, semi-parametric, multidimensional) statistics which are complementary and interlinked topics. It is essential to develop simultaneously statistical methods for the estimation and control methods for the optimization of the models.

3.2. Main research topics

- Stochastic modeling: Markov chain, Piecewise Deterministic Markov Processes (PDMP), Markov Decision Processes (MDP).
  The mathematical representation of complex systems is a preliminary step to our final goal corresponding to the optimization of its performance. For example, in order to optimize the predictive maintenance of a system, it is necessary to choose the adequate model for its representation. The step of modeling is crucial before any estimation or computation of quantities related to its optimization. For this we have to represent all the different states of the system and the behavior of the physical variables under each of these states. Moreover, we must also select the dynamic variables which have a potential effect on the physical variable and the quantities of interest. The team CQFD works on the theory of Piecewise Deterministic Markov Processes (PDMP's) and on Markov Decision Processes (MDP's). These two classes of systems form general families of controlled stochastic processes suitable for the modeling of sequential decision-making problems in the continuous-time (PDMPs) and discrete-time (MDP's) context. They appear in many fields such as engineering, computer science, economics, operations research and constitute powerful class of processes for the modeling of complex system.
- Estimation methods: estimation for PDMP; estimation in non- and semi parametric regression modeling.
  To the best of our knowledge, there does not exist any general theory for the problems of estimating parameters of PDMPs although there already exist a large number of tools for sub-classes of PDMPs such as point processes and marked point processes. However, to fill the gap between these specific models and the general class of PDMPs, new theoretical and mathematical developments will be on the agenda of the whole team. In the framework of non-parametric regression or quantile regression, we focus on kernel estimators or kernel local linear estimators for complete data or censored data. New strategies for estimating semi-parametric models via recursive estimation procedures have also received an increasing interest recently. The advantage of the recursive estimation approach is to take into account the successive arrivals of the information and to refine, step after step, the implemented estimation algorithms. These recursive methods do require restarting calculation of parameter estimation from scratch when new data are added to the base. The idea is to use only the previous estimations and the new data to refresh the estimation. The gain in time could be very interesting and there are many applications of such approaches.

- Dimension reduction: dimension-reduction via SIR and related methods, dimension-reduction via multidimensional and classification methods.
  Most of the dimension reduction approaches seek for lower dimensional subspaces minimizing the loss of some statistical information. This can be achieved in modeling framework or in exploratory data analysis context.
  In modeling framework we focus our attention on semi-parametric models in order to conjugate the advantages of parametric and nonparametric modeling. On the one hand, the parametric part of the model allows a suitable interpretation for the user. On the other hand, the functional part of the model offers a lot of flexibility. In this project, we are especially interested in the semi-parametric regression model $Y = f(X'\theta) + \varepsilon$, the unknown parameter θ belongs to $\mathbb{R}^p$ for a single index model, or is such that $\theta = [\theta_1, \cdots, \theta_d]$ (where each θ_k belongs to $\mathbb{R}^p$ and $d \leq p$ for a multiple indices model), the noise ε is a random error with unknown distribution, and the link function f is an unknown real valued function. Another way to see this model is the following: the variables X and Y are independent given $X'\theta$. In our semi-parametric framework, the main objectives are to estimate the parametric part θ as well as the nonparametric part which can be the link function f, the conditional distribution function of Y given X or the conditional quantile q_α. In order to estimate the dimension reduction parameter θ we focus on the Sliced Inverse Regression (SIR) method which has been introduced by Li [56] and Duan and Li [51]
  Methods of dimension reduction are also important tools in the field of data analysis, data mining and machine learning.They provide a way to understand and visualize the structure of complex data sets.Traditional methods among others are principal component analysis for quantitative variables or multiple component analysis for qualitative variables. New techniques have also been proposed to address these challenging tasks involving many irrelevant and redundant variables and often comparably few observation units. In this context, we focus on the problem of synthetic variables construction, whose goals include increasing the predictor performance and building more compact variables subsets. Clustering of variables is used for feature construction. The idea is to replace a group of "similar" variables by a cluster centroid, which becomes a feature. The most popular algorithms include K-means and hierarchical clustering. For a review, see, e.g., the textbook of Duda [52]
- Stochastic optimal control: optimal stopping, impulse control, continuous control, linear programming, singular perturbation, martingale problem.
  The first objective is to focus on the development of computational methods.
  - In the continuous-time context, stochastic control theory has from the numerical point of view, been mainly concerned with Stochastic Differential Equations (SDEs in short). From the practical and theoretical point of view, the numerical developments for this class of processes are extensive and largely complete. It capitalizes on the connection between SDEs and second order partial differential equations (PDEs in short) and the fact that the properties of the latter equations are very well understood. It is, however, hard to deny that the development of computational methods for the control of PDMPs has received little attention. One of the main reasons is that the role played by the familiar PDEs in the diffusion models is here played by certain systems of integro-differential equations for which there is not (and cannot be) a unified theory such as for PDEs as emphasized by M.H.A. Davis in his book. To the best knowledge of the team, there is only one attempt to tackle this difficult problem by O.L.V. Costa and M.H.A. Davis. The originality of our project consists in studying this unexplored area. It is very important to stress the fact that these numerical developments will give rise to a lot of theoretical issues such as type of approximations, convergence results, rates of convergence,....
  - Theory for MDP's has reached a rather high degree of maturity, although the classical tools such as value iteration, policy iteration and linear programming, and their various extensions, are not applicable in practice. We believe that the theoretical progress of MDP's must be in parallel with the corresponding numerical developments. Therefore, solving

MDP's numerically is an awkward and important problem both from the theoretical and practical point of view. In order to meet this challenge, the fields of neural networks, neuro-dynamic programming and approximate dynamic programming became recently an active area of research. Such methods found their roots in heuristic approaches, but theoretical results for convergence results are mainly obtained in the context of finite MDP's. Hence, an ambitious challenge is to investigate such numerical problems but for models with general state and action spaces. Our motivation is to develop theoretically consistent computational approaches for approximating optimal value functions and finding optimal policies.

Analysis of various problems arising in MDPs leads to a large variety of interesting mathematical problems. The second objective of the team is to study some theoretical aspects related to MDPs such as convex analytical methods and singular perturbation.

4. Application Domains

4.1. Application Domains

Our abilities in probability and statistics apply naturally to industry in particular in studies of dependability and safety.

An illustrative example which gathers all the topics of team is a collaboration started in May 2010 with Thales Optronique on the subject of *optimization of the maintenance of a digital camera equipped with HUMS* (Health Unit Monitoring Systems). This subject is very interesting for us because it combines many aspects of our project. Classification tools will be used to select significant variables as the first step in the modeling of a digital camera. The model will then be analysed and estimated in order to optimize the maintenance.

A second example concerns the optimization of the maintenance date for an aluminum metallic structure subject to corrosion. It is a structure of strategic ballistic missile that is stored in a nuclear submarine missile launcher in peace-time and inspected with a given periodicity. The requirement for security on this structure is very strong. The mechanical stress exerted on the structure depends on its thickness. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the break.

A third example is the minimization of the acoustic signature of a submarine. The submarine has to chose its trajectory in order to minimize at each time step its observability by a surface ship following an unknown random trajectory.

But the spectrum of applications of the topics of the team is larger and may concern many other fields. Indeed non parametric and semi-parametric regression methods can be used in biometry, econometrics or engineering for instance. Gene selection from microarray data and text categorization are two typical application domains of dimension reduction among others. We had for instance the opportunity via the scientific program PRIMEQUAL to work on air quality data and to use dimension reduction techniques as principal component analysis (PCA) or positive matrix factorization (PMF) for pollution sources identification and quantization.

5. Software

5.1. Package "ClustOfVar"

This R package is dedicated to cluster analysis of a set of variables. Variables can be quantitative, qualitative or a mixture of both. A new version of the package is available via the link http://cran.r-project.org/web/packages/ClustOfVar/index.html since novembre 2011. The new version improves the computational time of the "kmeansvar" function used for k-means type clustering of variables. This function is now able to deal with datasets of several thousands of variables like genomic data. The package is detailed in a paper submitted for publication [44]. It has been presented in several conferences [34], [35].

5.2. Package "PCAmixdata"

This package is dedicated to factorial analysis and rotation of quantitative data, qualitative data, or mixed data. The PCAMIX method, proposed in this package includes the ordinary principal component analysis (PCA) and multiple correspondence analysis (MCA) as special cases. Orthogonal varimax rotation of the principal components of PCAMIX is also implemented in this package. Theoretical and practical results about the new rotation algorithm available in the package is in revision for publication [45] and has been presented in [36].

6. New Results

6.1. On the Existence of Strict Optimal Controls for Constrained, Controlled Markov Processes in Continuous-Time

Participant: François Dufour.

Closedness and convexity conditions are identified under which optimal controls in the class of strict controls exist for a large class of stochastic processes under infinite-horizon discounted, long-term average, first exit, finite-horizon and discretionary stopping criteria in the presence of hard and/or soft constraints. The results are more general than results obtained by Haussmann and Lepeltier for a controlled diffusion under a mixed optimal-stopping/finite-horizon/first-exit criterion. The approach taken in this work is to utilize equivalent linear programming formulations of the control problems which provides a unified LP formulation for the problems. The conditions of Haussmann and Lepeltier are shown to imply the sufficient conditions of this paper when the process is a controlled diffusion. Simpler conditions are also identified for Markov chains, simple Markov jump processes, diffusions with jumps, regime-switching diffusions and solutions to Levy stochastic differential equations.

These results have been obtained in collaboration with Richard Stockbridge, Department of Mathematical Sciences, University of Wisconsin Milwaukee, USA. It has been accepted for publication in Stochastics [14]

6.2. Approximation of Markov Decision Processes with General State Space

Participant: François Dufour.

In this work, we deal with a discrete-time finite horizon Markov decision process with locally compact Borel state and action spaces, and possibly unbounded cost function. Based on Lipschitz continuity of the elements of the control model, we propose a state and action discretization procedure for approximating the optimal value function and an optimal policy of the original control model. We provide explicit bounds on the approximation errors. Our results are illustrated by a numerical application to a fisheries management problem.

These results have been obtained in collaboration with Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED, Madrid, Spain. It has been accepted for publication in Stochastics [13]

6.3. Asymmetry tests for Bifurcating Auto-Regressive Processes with missing data

Participants: Benoîte de Saporta, Anne Gégout-Petit.

Using the properties of the estimators studied in [20], we have constructed symmetry tests for bifurcating autoregressive processes (BAR) when some data are missing. BAR processes typically model cell division data. Each cell can be of one of two types *odd* or *even*. The goal of this work is to test asymmetry between odd and even cells in a single observed lineage. We have also derived asymmetry tests for the lineage itself, modeled by a two-type Galton-Watson process observed in a non standard scheme. We present applications on both simulated and real data.

This work is in collaboration with Laurence Marsalle of Lille 1 university. It will be soon submitted for publication.

6.4. Statistical study of asymmetry in cell lineage data

Participants: Benoîte de Saporta, Anne Gégout-Petit.

Simulation studies of the asymmetry in a single lineage data have shown a lack of power of the tests when the number of available generation is limited. This work proposes a rigorous methodology to study cell division data in the context of observation of several lineages. It generalizes [20]. We model the data by an asymmetric bifurcating autoregressive (BAR) process and take into account possibly missing data by modeling the genealogies with a two-type Galton Watson (GW) process. Our inference is based on several lineages, i.e. independent and identically distributed replicas of the coupled BAR and GW processes. We propose a least-squares estimator of the unknown parameters of the BAR process and an estimator of the parameters of the GW process, study their asymptotic properties and propose symmetry tests. Our results are applied on real data of Escherichia coli division.

This work is in collaboration with Laurence Marsalle of Lille 1 university. It will be soon submitted for publication.

6.5. Estimation of the jump rate of a PDMP

Participants: Romain Azaïs, François Dufour, Anne Gégout-Petit.

We estimate the jump rate of PDMP. We suppose the flow given by physics laws and we want to make some inference on λ. ϕ being deterministic, the problem can be rewritten as a problem of estimation of the rate $\lambda(z,t)$ with $z \in E$ with E an open set of a separable metric space. We have an ergodicity assumption on the observed PDMP and the asymptotic is in the time of observation of the process.

We distinguish three cases :

1. E is finite. In this case, we easily estimate each of the cumulated risk functions $\Lambda(z,t) = \exp(-\int_0^t \lambda(z,s)ds)$ corresponding to each of $z \in E$ by a Nelson Aalen estimator. The results is based on the decomposition in semi-martingale of the following counting process in an appropriate filtration:

$$\forall t \geq 0,\ N_n(z,t) = \sum_{i=0}^{n-1} \mathbf{1}_{\{S_{i+1} \leq t\}} \mathbf{1}_{\{Z_i = z\}}, \tag{1}$$

   We obtain the estimator of the rate $\lambda(z,t)$ by smoothing of the estimator of Λ.

2. E is an open set of a general separable metric space but the transition measure Q does not depend on the time spent in the current regime. In this case, we suppose the rate $\lambda(z,t)$ Lipschitz and the process ergodic with a stationary law denoted by ν. We first construct an estimation of the cumulated rate knowing that z belongs to a set A such that $\nu(A) > 0$ by :

$$\widehat{L}_n(A,t) = \sum_{i=0}^{n-1} \frac{1}{Y_n(A, S_{i+1})} \mathbf{1}_{\{S_{i+1} \leq t\}} \mathbf{1}_{\{Z_i \in A\}} \qquad \text{with} \qquad Y_n(A,t) = \sum_{i=0}^{n-1} \mathbf{1}_{\{S_{i+1} \geq t\}} \mathbf{1}_{\{Z_i \in A\}}. \tag{2}$$

We show the consistence of the estimator. Smoothing $\widehat{L}_n(A,t)$ and using a fine partition of E allow us to obtain an uniform result for the approximation of the rate $\lambda(z,t)$, in some sense in t and z.

3. E is an open set of a general separable metric space and the transition measure Q depends on the time spent in the current regime. Here, we loose some conditional independence between the S_i's and the whole set of the locations of the jump $\{Z_1, ..., Z_n\}$. We have to make a detour for the estimation of the law of the time S_{k+1} knowing the current Z_k by the the law S_{k+1} knowing (Z_k, Z_{k+1}). The method gives an estimation of the conditional density of S_{k+1} given Z_k.

We have made simulation studies that give expected results. A R package for this estimation method is in progress.

This work is a part of the PhD Thesis of R. Azaïs founded by the ANR Fautocoes. R. Azaïs has presented a part of this work at "Rencontres des Jeunes Statisticiens" in 2011 September [28]. The work will be soon submitted to a international peer-reviewed journal for publication.

6.6. Detection of a damaged operating mode of optronic equipment using Hidden Markov Model

Participants: Camille Baysse, Anne Gégout-Petit, Jérôme Saracco.

As part of optimisation of the reliability, Thales Optronics now includes systems that examine the state of its equipment. This function is performed by HUMS (Health & Usage Monitoring System). The aim is to implement a program based on these observations that can determine the lifetime of this optronic equipment and optimize its maintenance.

Our study focuses on a simple example of HUMS. As part of our research, we are interested in a variable called "time-to cold" noted TMF, which reflects the state of the system. Using these informations about this variable, we seek to detect as soon as possible a damaged state and propose a maintenance before failure. This would allow the Thales Optronics company to improve its maintenance system and achieve many economies.

For this we use a hidden Markov model. The state of our system at time t is then modeled by a continuous time Markov chain X(t) with three states : stable, damaged and failure. However we do not observe directly this chain but indirectly through the TMF, a noisy function of this chain. Thanks to filtering equations, we obtained results on the probability that an equipment is in a damaged state at time t, knowing the history of the TMF until this moment. We have subsequently studied the method on simulated data., before applying these results on the analysis of our real data and we have checked that the results are consistent with the reality. This work will be used by Thales Optronique for the optimization of the maintenance.

This work is a part of the CIFRE PhD Thesis of Camille Baysse founded by Thales Optronic. It is the object of a technical report [48] and was presented in [29] in an internal Thales seminar.

6.7. Optimal quantization applied to Sliced Inverse Regression

Participants: Romain Azaïs, François Dufour, Anne Gégout-Petit, Jérôme Saracco.

We tackle the well known Slice Inverse Regression (SIR) method for a semiparametric regression model involving a quantitative variable X and including a dimension reduction of X via a parameter β. The response variable Y is real. Our goal is to estimate β and to predict the response variable conditionally to X. We adapt SIR method using optimal quantization [57] in the first time only for the independent variable X for the estimation of β. In a second time, we quantize the variable $(\widehat{\beta}_n, Y)$ in order to propose a discrete conditional law of Y given $X = x$. We show the convergence of the estimator of β and of the conditional law. Simulation studies show the numerical qualities of our estimates. This work is the object of a publication in Journal of Statistical Planning and Inference [15] and was presented in a national conference [23]

6.8. Multivariate Analysis for the detection of the effect of a treatment

Participant: Anne Gégout-Petit.

The aim of this work is to give some statistical rules to determine if a patient is meeting a given treatment (a BD here). The criterium commonly used to determine if a patient is meeting a BD treatment is based only on one physiological parameter : if this parameter increases, the patient is meeting. But now, many physiological parameters are measured in routine and physiologists '"ont le sentiment" that a patient could have a global amelioration of his health state due to the treatment without an increase of the single used parameter.

Using the measures of six variables before and after the treatment, the expected value of this variable under the hypothesis of good health, we first propose indices of amelioration. Using standard multivariate analysis techniques, we first study the correlation between these indices. We use classification in order to constitute groups of patients whose present homogeneous treatment responses. The method used on a cohort of 100 subjects gives three groups : in the first one, the mean of the indices is near zero, the treatment has no effect. In the second one, subjects present significant amelioration regarding two of the indices but not the indices related to the commonly used one. The last group show an amelioration for all the indices.

We have proposed criteria to discriminate the three groups. These criteria where used on a validation cohort to constitute three groups. Their global characteristics were the same as in the original cohort and it comforts the validity of the method. This work will be used by physiologists to propose new criteria for the measure of the effect of a BD treatment. It is in collaboration with physiologists from Bordeaux and Nantes universities and is the object of a paper that will be soon submitted in a international peer-reviewed journal in the domain of pneumology.

6.9. Numerical computation of expectations of PDMP's

Participants: Adrien Brandejsky, Benoîte de Saporta, François Dufour.

This work concerns the computation of expectations of functionals of piecewise deterministic Markov processes (PDMP's). We propose a numerical scheme to approximate such expectations, analyze the convergence of our scheme and derive bounds for the convergence rate. More precisely, we are interested in the approximation of expectations of the form

$$E_x\left[\int_0^{T_N} l(X_t)dt + \sum_{j=1}^{N} c(X_{T_j^-})I_{\{X_{T_j^-}\in\partial E\}}\right]$$

where $(X_t)_{t\geq 0}$ is a PDMP, (T_j) are its jump times and l and c are some non negative, real-valued, bounded functions. Such expectations are discussed by M.H.A. Davis in [50], chapter 3. They often appear as *cost* or *reward* functions in optimization problems. The first term is referred to as the running cost while the second may be called the boundary jump cost. Besides, they are quite general since M.H.A. Davis shows how a *wide variety of apparently different functionals* can be obtained from the above specific form. For example, this wide variety includes quantities such as a mean exit time and even, for any fixed $t \geq 0$, the distribution of X_t (i.e. $E_x[I_F(X_t)]$ where F is a measurable set).

There are surprisingly few works in the literature devoted to the actual computation of such expectations, using other means than direct Monte Carlo simulations. M.H.A Davis showed that these expectations satisfy integro-differential equations. However, the set of partial differential equations that is obtained is unusual. Roughly speaking, these differential equations are basically transport equations with a non-constant velocity and they are coupled by the boundary conditions and by some integral terms involving kernels that are derived from the properties of the underlying stochastic process. This approach is currently under study in this project by LATP. The main difficulty comes from the fact that the domains on which the equations have to be solved vary from one equation to another making their numerical resolution highly problem specific. Another similar approach has been recently investigated in [49], [54]. It is based on a discretization of the Chapman Kolmogorov equations satisfied by the distribution of the process $(X_t)_{t\geq 0}$. The authors propose an approximation of such expectations based on finite volume methods. Unfortunately, their method is only valid if there are no jumps at the boundary.

Our approach is completely different and does not rely on differential equations, but on the fact that such expectations can be computed by iterating an integral operator G. This operator only involves the embedded Markov chain $(Z_n, S_n)_{n\in N}$ and conditional expectations. It is therefore natural to propose a computational method based on the quantization of this Markov chain, following the same idea as [8]. We also addressed two important aspects that had not been investigated in [8]. The first one consists in allowing c and l to be time depending functions, although still Lipschitz continuous, so that we may compute expectations of the form

$$E_x\left[\int_0^{T_N} l(X_t,t)dt + \sum_{j=1}^{N} c(X_{T_j^-},T_j)I_{\{X_{T_j^-}\in\partial E\}}\right].$$

This important generalization has huge applicative consequences. For instance, it allows discounted *cost* or *reward* functions such as $l(x,t) = e^{-\delta t}l(x)$ and $c(x,t) = e^{-\delta t}c(x)$ where δ is some interest rate. To compute the above expectation, our strategy consists in considering, as it is suggested by M.H.A. Davis in [50], the time augmented process $\widetilde{X}_t = (X_t, t)$.

The second important generalization is to consider the deterministic time horizon problem. Indeed, it seems crucial, regarding to the applications, to be able to approximate

$$\begin{aligned} & E_x\left[\int_0^{t_f} l(X_t,t)dt + \sum_{T_j\le t_j} c(X_{T_j^-},T_j)I_{\{X_{T_j^-}\in\partial E\}}\right] \\ = \quad & E_x\left[\int_0^{+\infty} l(X_t,t)I_{\{t\le t_f\}}dt + \sum_{j=1}^{+\infty} c(X_{T_j^-},T_j)I_{\{X_{T_j^-}\in\partial E\}}I_{\{T_j\le t_f\}}\right] \end{aligned}$$

for some fixed $t_f > 0$ regardless of how many jumps occur before this deterministic time. To compute this quantity, we start by choosing a time N such that $P(T_N < t_f)$ be small so that the previous expectation boils down to $E_x\left[\int_0^{T_N} l(X_t,t)I_{\{t\le t_f\}}dt + \sum_{j=1}^{N} c(X_{T_j^-},T_j)I_{\{X_{T_j^-}\in\partial E\}}I_{\{T_j\le t_f\}}\right]$. We deal with the two indicator functions in two different ways. On the one hand, we prove that it is possible to relax the regularity condition on the running cost function so that our algorithm still converges in spite of the first indicator function. On the other hand, since the same reasoning cannot be applied to the indicator function within the boundary jump cost term, we bound it between two Lipschitz continuous functions. This provides bounds for the expectation of the deterministic time horizon functional.

An important advantage of our method is that it is flexible. Indeed, as pointed out in [47], a quantization based method is *obstacle free* which means, in our case, that it produces, once and for all, a discretization of the process independently of the functions l and c since the quantization grids merely depend on the dynamics of the process. They are only computed once, stored off-line and may therefore serve many purposes. Once they have been obtained, we are able to approximate very easily and quickly any of the expectations described earlier. This flexibility is definitely an important advantage of our scheme over standard methods such as Monte-Carlo simulations since, with such methods, we would have to run the whole algorithm for each expectation we want to compute.

The theoretical part of this work with rigorous proofs is under review for an international peer-reviewed journal [10]. F. Dufour presented this work in an invited session at an international conference [22].

6.10. Optimal stopping under partial observation

Participants: Adrien Brandejsky, Benoîte de Saporta, François Dufour.

In continuation of our work on optimal stopping for Piecewise deterministic Markov processes (PDMP's) [8], we have started investigating the optimal stopping problem when the process is only partially observed. We supposed that the jump times of the process are observed, but the post jump locations are observed through a noise.

The first step is to rewrite the optimal stopping problem for the partially observed PDMP as a totally observed stopping problem for a new Markov chain, obtained by filtering the observation process. Then, one has to study precisely this filter, which is non standard due to the possible jumps of the process. The next step is to derive the dynamic programming equation adapted to our framework. Finally, we propose a numerical method based on quantization to approximate the value function and ϵ-stopping times. Track is also kept of the error bounds all through our numerical procedure.

This work is still in progress and should be submitted to an international peer-reviewed journal shortly.

6.11. Efficient simulation of the availability of a feedwater control system

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

In the reliability modeling of complex control systems, classical methodologies such as even-trees/fault-trees or Petri nets may not represent adequately the dynamic interactions existing between the physical processes (modeled by continuous variables) and the functional and dysfunctional behavior of its components (modeled by discrete variables). This paper proposes a framework for modeling and simulation of a water level control system in the steam generator (SG) in the secondary circuit of a nuclear power plant. A similar benchmark system was described by the U.S. Nuclear Regulatory Commission [46] to compare two approaches to dynamic reliability: DFM (Dynamic Flowgraph Methodology) and Markov/CCMT (Cell-to-Cell Mapping Technique). But the report released by the NRC is not sufficient to reconstruct a realistic model. We have developed a complete benchmark case. The behavioral model of SG is obtained from a linearized model published in 2000 by EDF [55]. Detailed description of the components, failure modes and control laws of the principal components is presented. For modeling the system, we use the piecewise deterministic Markov processes (PDP) framework [50] and for implementation we chose Simulink associated with Stateflow. PDP's offer a very general modeling framework to deal with dynamic reliability problems; Simulink is a good tool to simulate non linear differential equations and their controller, while Stateflow implementation is appropriate for finite state machine descriptions of different components.

In our benchmark system, four physical processes are considered: feedwater flowrate , steam flow, narrow range water level and wide range water level. A PID controller is used to maintain the water level within limits of set-points. The system is composed of seven components: 1 passive system representing vapor transport system, 3 extraction pumps, 2 feeding turbopumps, and 1 waterflow regulation valve. The functional and dysfunctional behaviors and the failure rates of each component are based on operational experience.

We show that PDP modeling is suitable for dynamic reliability analysis, and that Simulink associated with Stateflow provides an interactive simulator (graphical block diagramming), which makes the simulator scalable. This work is submitted for presentation to an international conference in 2012.

6.12. Othogonal Rotation in PCAMIX

Participants: Marie Chavent, Jérôme Saracco.

The aim of this work is to propose an efficient algorithm for rotation in PCAMIX, a principal component method for a mixture of qualitative and quantitative variables. We give a new presentation of PCAMIX where the principal components and the squared loadings are obtained from a Singular Value Decomposition. The loadings of the quantitative variables and the principal coordinates of the categories of the qualitative variables are also obtained directly. In this context, we propose a computationaly efficient procedure for varimax rotation in PCAMIX and a direct solution for the optimal angle of rotation. A simulation study shows the good computational behavior of the proposed algorithm. An application on a real data set illustrates the interest of using rotation in MCA. All source codes are available in the R package "PCAmixdata". This work is in revision for publication [45] and has been presented in [36].

6.13. An adaptive SIR method for block-wise evolving data streams

Participants: Marie Chavent, Jérôme Saracco.

In this work, we consider block-wise evolving data streams. When a semi-parametric regression model involving a common dimension reduction direction β is assumed for each block, we propose an adaptive SIR (for sliced inverse regression) estimator of β. This estimator is faster than usual SIR applied to the union of all the blocks, both from computational complexity and running time points of view. We show the consistency of our estimator at the root-n rate and its asymptotic normality. We also propose an extension of this method to multiple indices model. In simulation studies, we illustrate the good numerical behavior of our estimator. We also provide a graphical tool in order to detect if there exists a drift of the dimension reduction direction or some aberrant blocks of data. We illustrate our approach with various scenarios. We apply this approach on the following real data problem.

As an illustration, we consider a nonlinear inverse problem in remote sensing. The goal is to estimate the physical properties of surface materials on the planet Mars from hyperspectral data. The method is based on the estimation of the functional relationship between some physical parameters Y and observed spectra X. For this purpose, a database of synthetic spectra is generated by a physical radiative transfer model. We propose to reduce the high dimension of spectra ($p = 352$ wavelengths) with a regularized version of SIR. The need to regularize SIR in very high dimensions is well-known. In practice, the database of synthetic spectra may be so large that it cannot be stored in a computer memory. Thus, a stream of smaller sub-databases is generated and we apply our "SIR datastream" approach to this context.

This work will be submitted for publication very soon and it has been presented in the international conference [31].

6.14. Classification of EEG data by evolutionary algorithm for the study of vigilance states

Participants: Marie Chavent, Laurent Vézard.

The objective of this work [42] is to predict the state of vigilance of an individual from the study of its brain activity (EEG signals). The variable to predict is binary (alertness "normal" or "relaxed"). EEG of 44 participants in both states (88 records) were collected with a helmet with 58 electrodes. After a pretreatment step and data validation, a test called "test slope" was chosen. The usual methods of supervised classification (k nearest neighbors, binary classification trees, random forests, and discriminant sparse PLS) were used to provide predictions of the state of participants. The test was then refined using a genetic algorithm, which has built a reliable model (average true classification rate by using CART equal to $86.68 \pm 1.87\%$) and to select an electrode from the initial 58. This work is in collaboration with Pierrick Legrand (EPI Alea) and Frédérique Faïta (EA 487 "Sciences cognitives et facteur humain"

6.15. Comparison of Kernel Density Estimators on Environmental Data with Assumption on Number of Modes

Participants: Jérôme Saracco, Raphaël Coudret.

In this work, we consider valvometric data samples, that is measurements of distances between the two parts of the shell of bivalves. The movements of a few oysters are monitored in different places (like Arcachon Bay or Santander port) by a laboratory called Environnements et Paléoenvironnements Océaniques et Continentaux (EPOC). The aim of these experiments is to determine water quality via the observation of the bivalves behavior. Previous related works have been published on this subject. EPOC team studied this animals in a copper pollution context via investigations using kernels methods. In our study, we consider each day the density of the distance between the two parts of the shell. This density is reasonably assumed to have 2 modes, the first one corresponds to a close status of the shell and the second one to an open status. The study of the evolution of this density along the time provides informations on bivalves behavior. We provide theoretical

results on our bandwidth choice with a bounded support kernel and we give a procedure to determine this bandwidth. We also provide asymptotic results for a density kernel estimator with a kernel that has it support on $\mathfrak{R}$. We present an extensive simulation study in order to compare numerical performances of various density estimators based or not on the two modes assumption. From the results obtained from the simulated data, we derive the suitable estimator for our real data application. This work will be submitted for publication very soon and it has been presented in the international conference [37].

7. Contracts and Grants with Industry

7.1. Astrium

Participants: Romain Azaïs, Adrien Brandejsky, Benoîte de Saporta, François Dufour, Anne Gégout-Petit, Huilong Zhang.

The goal of this project is to propose models for fatigue of structure and to study an approach to evaluate the probability of occurrence of events defined by the crossing of a threshold. In this context, Astrium funds the PhD Thesis of Adrien Brandejsky since september 2009 and is a partner of ANR Fautocoes.

7.2. Thales Optronique

Participants: Camille Baysse, Benoîte de Saporta, François Dufour, Anne Gégout-Petit, Jérôme Saracco.

The goal of the project is the optimization of the maintenance of a on board system with a HUMS (Health Unit Monitoring Systems). The collaboration is the subject of the PhD of Camille Baysse (CIFRE) on this subject.

7.3. DCNS

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

In september 2010, an industrial collaboration started with DCNS on the application of Markov Decision Processes to optimal stochastic control of a submarine to maximize the acoustic signature of a target vessel. In 2011, we extended our previous results to multiple target vessels. This work was presented at the INRIA seminar *Unithé ou café* and gave rise to a new technical report *Contrôle optimal stochastique appliqué à l'optimisation de trajectoire, cas multicible* [53].

7.4. EDF Approdyn

Participants: Benoîte de Saporta, François Dufour, Huilong Zhang.

The objective of this project is develop new methodologies for studying the dynamic reliability of controlled systems used in the critical area of power generation and process industries. We work on a benchmark of steam generator with four physical processes: feedwater flowrate , steam flow, narrow range water level and wide range water level. A PID controller is used to maintain the water level within limits of set-points. The system is composed of seven components: 1 passive system representing vapor transport system, 3 extraction pumps, 2 feeding turbopumps, and 1 waterflow regulation valve. This work was presented as a poster at the 3SGS GIS workshop in Valenciennes in 2011.

7.5. Lyre (Lyonnaise Recherche)

Participant: Jérôme Saracco.

A project of collaboration started in december 2011 between LyRE (which is a research laboratory of Lyonnaise des Eaux, subsidiary company of Suez environnement) and INRIA team CQFD. One aim of this project is to develop tools for detection of leaks in water network based on a sampling study of continuous time monitoring of individual consumptions. A PhD thesis (2011-2014) is part of this project. The whole project also involves CEMAGREF and IMB (Institut de Mathématiques de Bordeaux). This work is in collaboration with Vincent Couallier (IMB).

7.6. EDF

Participant: Jérôme Saracco.

A new contract was signed in 2011 between EDF R & D team ICAME and INRIA teams CQFD and ALEA. This contract deals with modeling and forecasting of short term electricity loads for private client of EDF. Its purpose is to propose accurate short-term forecasts (each hour) for particular clients by using past electricity loads and temperature loads. Parametric models (such as ARMAX or SARIMAX) are semiparametric models (based on sliced inverse regression) have been studied and evaluated on the data. The total value of the contract is 20 000 euros.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Marie Chavent participates to a project financed by the Région Aquitaine for three years (2010-2013), named *PSI : Etude des interactions états psychophysiologiques et musique* including the PHD-grant of Laurent Vezard. The subject of this PHD, co-directed by M. Chavent, F. Faita and P. Legrand from Project-Team ALEA, is *Dimension reduction in the context of supervised learning. Applications to the electrical brain activity study*.

8.2. National Initiatives

8.2.1. ANR FAUTOCOES

The goal of the project "FAUTOCOES" (number ANR-09-SEGI-004) of the ARPEGE program of the French National Agency of Research (ANR) can be described as follows. Today, complex technological processes must maintain an acceptable behavior in the event of random structural perturbations, such as failures or component degradation. Aerospace engineering provides numerous examples of such situations: an aircraft has to pursue its mission even if some gyroscopes are out of order, a space shuttle has to succeed in its re-entry trip with a failed on-board computer. Failed or degraded operating modes are parts of an embedded system history and should therefore be accounted for during the control synthesis.

These few basic examples show that complex systems like embedded systems are inherently vulnerable to failure of components and their reliability has to be improved through fault-tolerant control. Embedded systems require mathematical representations which are in essence dynamic, multi-model and stochastic. This increasing complexity poses a genuine scientific challenge:

- to model explicitly and realistically the dynamical interactions existing between the physical state variables defining the system: pressure, temperature, flow rate, intensity, etc, and the functional and dysfunctional behavior of its components;
- to estimate the performance of the system through the evaluation of reliability indexes such as availability, quality, and safety;
- to optimize the control to prevent system failures, as well as to maintain the system function when a failure has occurred.

Our aim is to meet the previously mentioned challenge by using the framework of piecewise deterministic Markov processes (PDMP's in short) with an emphasis on probabilistic and deterministic numerical methods. More precisely, our objectives are

- to use the framework of piecewise deterministic Markov processes to model complex physical systems and phenomena;
- to compute expectations of functionals of the process in order to evaluate the performance of the system;
- to develop theoretical and numerical control tools for PDMP's to optimize the performance and/or to maintain system function when a failure has occurred.

More details are available at http://fautocoes.bordeaux.inria.fr/.

8.2.2. ANR ADAPTEAU

The ANR project ADAPTEAU has been obtained for the period 2012-2016 and will start in january 2012.

ADAPTEAU aims to contribute to the analysis and management of global change impacts and adaptation patterns in River-Estuarine Environments (REEs) by interpreting the scientific challenges associated with climate change in terms of: i) scale mismatches; ii) uncertainty and cognitive biases between social actors; iii) interdisciplinary dialogue on the "adaptation" concept; iv) critical insights on adaptive governance and actions, v) understanding the diversity of professional, social and economic practices vis-à-vis global change. The project aims to build an integrative and interdisciplinary framework involving biophysical and social sciences, as well as stakeholders and civil society partners. The main objective is to identify adaptive strategies able to face the stakes of global change in REEs, on the basis of what we call 'innovative adaptation options'.

We consider the adaptation of Social-Ecological Systems (SES) through the expected variations of the hydrological regimes (floods / low-flow) of the Garonne-Gironde REE—a salient issue in SW France, yet with a high potential for genericity The ADAPTEAU project will be organised as follows:

- Achieve and confront socio-economic and environmental assessments of expected CC impacts on the Garonne-Gironde river-estuarine continuum (task 1);
- Identify the emerging 'innovative adaptation options' endorsed by various social, economic, political actors of the territory (depolderisation, 'room for rivers' strategies, changes in economic activities, agricultural systems or social practices), then test their environmental, economic and social robustness through a selected subset (task 2);
- Scientists, representatives from administrators and civil society collaborate to build adaptation scenarios, and discuss them in pluralistic arenas in order to evaluate their social and economic feasibility, as well as the most appropriate governance modes (task 3).
- Disseminate the adaptation strategies to academics and managers, as well as to the broader society (task 4).

  The expected results are the definition and diffusion of new regional-scale reference frameworks for the discussion of adaptation scenarios in REE and other SESs, as well as action guidelines to better address climate change stakes.

THE CQFD team will work on tasks 1 and 3.

8.3. International Initiatives

8.3.1. Visits of International Scientists

+ Oswaldo Luiz do Valle Costa, Escola Politécnica da USP, Universidade de Sao Paulo, has visited F. Dufour (2 weeks).
+ F. Dufour has visited Oswaldo Luiz do Valle Costa, Escola Politécnica da USP, Universidade de Sao Paulo (2 weeks).
+ F. Dufour has visited Alexey Piunovskiy, University of Liverpool (2 weeks).
+ Alexey Piunovskiy, University of Liverpool, has visited F. Dufour (2 weeks).
+ Tomas Prieto-Rumeau, University of Madrid, has visited F. Dufour (2 weeks).
+ Richard Stockbridge, University of Wisconsin, Milwaukee, has visited F. Dufour (2 weeks).
+ F. Dufour has visited Richard Stockbridge, University of Wisconsin, Milwaukeel (2 weeks).

9. Dissemination

9.1. Editorial activities

M. Chavent is member of the scientific committee of SFC'11.

F. Dufour is associate editor of the journal: SIAM Journal of Control and Optimization since 2009.

All the member of the team are regular reviewers for the most important journals in applied probability and statistics.

9.2. Scientific responsibilities

M. Chavent has been elected to National Council of the Universities (CNU).

B. de Saporta is in charge of the "Tache 3" of the ANR project FAUTOCOES.

F. Dufour is the leader of the ANR project FAUTOCOES.

F. Dufour is member of the IFAC Technical Committee TC 1.4 Stochastic Systems, term Period 2008-2011.

9.3. Organization of workshops and conferences

The team CQFD will organize the first french-speaking meeting on the software R in July 2012.

9.4. Administration of the universities and research institutes

M. Chavent is co-director of the cursus *Modélisation Statistique et Sochastique* of the master MIMSE *Ingénierie Mathématique, Statistique et Economique* of the University of Bordeaux.

B. de Saporta is president of the "Congress and Colloquium" commission of the INRIA Bordeaux Sud-Ouest.

B. de Saporta is a member of the editorial board of SO News, the journal of INRIA Bordeaux Sud-Ouest.

B. de Saporta is correspondant of the cursus *Ingénierie Economique* of the master MIMSE *Ingénierie Mathématique, Statistique et Economique* of the University of Bordeaux.

B. de Saporta is in charge of the seminar of the team "Statistics and Probability" of the Institute of Mathematics of Bordeaux (IMB).

B. de Saporta is an elected (deputy) member of the CNU 26.

F. Dufour is member of the scientific council of the engineering school ENSEIRB-MATMECA.

F. Dufour is now vice-president of the INRIA Project Comity.

J. Saracco is member of the commission INRIA "Jeunes Chercheurs".

A. Gégout-Petit is member of the CEVU (Conseil des Etudes et de la Vie Universitaire) of the Bordeaux Segalen University.

A. Gégout-Petit is member of the Council of the Institut de Mathématiques de Bordeaux.

J. Saracco is the leader of the team "Statistics and Probability" of the Institute of Mathematics of Bordeaux (IMB). J. Saracco is an elected (deputy) member of the CNU 26.

H. Zhang is director of the cursus *Ingénierie Mathématique* of the Licence de Mathématiques of the University of Bordeaux.

9.5. Administration of the learned societies

M. Chavent was an elected member of the administration council of the SFdS until June 2011.

B. de Saporta belongs to the board of SMAI-MAS group. She was webmaster of the website until september 2011.

A. Gégout-Petit is an elected member of the administration council of the SFdS (Société Française de Statistique); she was vice-secretary of the SFDS until may 2010 and is now general secretary of the association.

A. Gégout-Petit is in the board team of the web-domain *emath.fr*. In this function, she manages the project "carte des masters" a web site which gathers together the informations on all the french masters in mathematics.

A. Gégout-Petit is in the organization committee of the "1er Forum Emploi Mathématiques" that will take place in Paris in January 2012

9.6. Promotion, dissemination of the science

B de Saporta is a member of the "Cellule Grand Public" of the SMAI.

A. Gégout-Petit is in charge for the promotion of "Licence MASS" (Applied mathematics degree) of the University of Bordeaux 2 to the secondary school pupils.

9.7. Teaching

Marie Chavent :

Licence : Statistique descriptive, 36 ETD ,L1, university Bordeaux Segalen, France

Master : Analyse des données 1, 43 ETD, niveau M1, university Bordeaux Segalen, France

Master : Modèle de régression, 29 ETD, niveau M1, university Bordeaux Segalen, France

Master : Logiciels de statistique, 12 ETD, niveau M1, university Bordeaux Segalen, France

Master : Analyse des données 2, 25 ETD, niveau M2, university Bordeaux Segalen, France

Master : Projet Informatique, 10 ETD, niveau M2, university Bordeaux Segalen, France

Master : Scoring, 21 ETD, niveau M2, university Montesquieu Bordeaux 4, France

PhD : Laurent Vezard, "Classification de signaux EEG et synthèse de paramètres musicaux par algorithme évolutionnaire", University of Bordeaux 1 (co-supervised with P. Legrand).

Benoîte de Saporta :

Licence: Mathématiques pour l'économie, 45h ETD, L3, Univ. Montesquieu Bordeaux IV, France.

Master: Processus Aléatoires en finance, 30h ETD, M1, Univ. Montesquieu Bordeaux IV, France,

Master: Processus à espace d'états discret, 25h ETD, M2, Univ. Segalen Bordeaux 2, France,

Master: Finance en temps discret, 29h ETD, M2, Univ. Bordeaux 1, France,

Master: Finance en temps continu, 10h ETD, M2, Univ. Montesquieu Bordeaux IV, France.

PhD : Adrien Brandejsky, Contribution à l'étude des processus Markoviens déterministes par morceaux, University of Bordeaux 1 (co-supervised with F. Dufour and C. Elegbede).

Anne Gegout-Petit

Licence Mathématiques Appliquées et Sciences Sociales, L3, Etudes de cas en statistique, 30h ETD

Licence Mathématiques Appliquées et Sciences Sociales, Econométrie et séries chronologiques, 36h ETD, Univ. Bordeaux Segalen, France.

Master Modélisation, Ingénierie Mathématique, Statistique et Economique M1 : Analyse de variance, 25h ETD.

PhD: Romain Azaïs, Méthodes d'estimation pour les Processus Markoviens Déterministes par Morceaux, University of Bordeaux 1 (co-supervised with F. Dufour).

PhD: Camille Baysse, Analyse et optimisation de la fiabilité d'un équipement opto-électronique équipé de HUMS, University of Bordeaux 1 (co-supervised with J.Saracco).

François Dufour :

Licence : Probabilités et statistiques, 16 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France. Probabilités , 10,6 heures, niveau L3, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

Master : Méthodes numériques pour la fiabilité, 24 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France. Probabilités, 20 heures, niveau M1, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France.

PhD : Romain Azaïs, Méthodes d'estimation pour les Processus Markoviens Déterministes par Morceaux, University of Bordeaux 1 (co-supervised with A. Gégout-Petit).

PhD : Adrien Brandejsky, Contribution à l'étude des processus Markoviens déterministes par morceaux, University of Bordeaux 1 (co-supervised with B. de Saporta and C. Elegbede)

Jérôme Saracco teaches mathematics, statistics and statistical modeling at the engineering school ENSC (Ecole Nationale Supérieure de Cognitique, a school of Cognitive Science Engineering) which is one of the six engineering schools of the Bordeaux Institute of Technology (IPB). He also teaches sampling techniques and experimental designs in the cursus "Statistical and stochastic modeling" of the Master "Ingénierie Mathématique, Statistique et Economique" at the University of Bordeaux. More precisely,

Licence: Descriptive statistics, 10.5h, L3, First year of ENSC, France

Licence: Mathematical statistics, 20h, L3, First year of ENSC, France

Licence: Data analysis (multidimensional statistics), 20h, L3, First year of ENSC, France

Licence: Mathematics (complement of linear algebra), 20h, L3, First year of ENSC, France

Master: Mathematics (complement of linear algebra and analysis), 20h, M1, First year of ENSC, France

Master: Statistical modeling, 20h, M1, Second year of ENSC, France

Master: training project, 20h, M1, Second year of ENSC, France

Master: Sampling techniques and experimental designs, 25h, M2, Master "Ingénierie Mathématique, Statistique et Economique", the University of Bordeaux, France

PhD: Camille Baysse, Analyse et optimisation de la fiabilité d'un équipement opto-électronique équipé de HUMS, University of Bordeaux 1 (co-supervised with A. Gégout-Petit)

PhD:Raphaël Coudret, Modélisation statistique de données acquises à haute fréquence : application en environnement et génétique, University of Bordeaux 1 (co-supervised with G.Durrieu, Université de Bretagne Sud)

Huilong Zhang

Licence Mathématiques, L1, Mathématiques de Base 89 heure, Université Bordeaux 1, France

Master Modélisation, Ingénierie Mathématique, Statistique et Economique M1: Outils de simulation I, 29 heures. Files d'attente, 29 heures, Projet tutoré 10 heures, Université Bordeaux 1, France

Licence : Intégration et Probabilités, 26 heures, Institut Polytechnique de Bordeaux, école ENSEIRB-MATMECA, France

10. Bibliography

Major publications by the team in recent years

[1] M. CHAVENT, B. LIQUET, J. SARACCO. *A semiparametric approach for a multivariate sample selection model*, in "Statist. Sinica", 2010, vol. 20, n^{o} 2, p. 513–536.

[2] O. COSTA, F. DUFOUR. *Stability and ergodicity of piecewise deterministic Markov processes*, in "SIAM J. Control Optim.", 2008, vol. 47, n° 2, p. 1053–1077.

[3] O. COSTA, F. DUFOUR. *The Vanishing Discount Approach for the Average Continuous Control of Piecewise Deterministic Markov Processes*, in "Journal of Applied Probability", 2009, vol. 46, n° 4, p. 1157-1183.

[4] F. DUFOUR, A. PIUNOVSKIY. *Multi-objective stopping problem for discrete-time Markov processes*, in "Journal of Applied Probability", 2010, vol. 47, n° 4, p. 947-966.

[5] A. GANNOUN, J. SARACCO, A. YUAN, G. E. BONNEY. *Non-parametric quantile regression with censored data*, in "Scand. J. Statist.", 2005, vol. 32, n° 4, p. 527–550.

[6] J. SARACCO. *Asymptotics for pooled marginal slicing estimator based on* SIR_α *approach*, in "J. Multivariate Anal.", 2005, vol. 96, n° 1, p. 117–135.

[7] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*, in "Electronic Journal of Statistics", 2011, vol. 5, p. 1313-1353 [*DOI :* 10.1214/11-EJS643], http://hal.inria.fr/hal-00545447/en.

[8] B. DE SAPORTA, F. DUFOUR, K. GONZALEZ. *Numerical method for optimal stopping of piecewise deterministic Markov processes*, in "The Annals of Applied Probability", 2010, vol. 20, n° 5, p. 1607-1637.

Publications of the year

Articles in International Peer-Reviewed Journal

[9] B. BERCU, T. MONG NGOC NGUYEN, J. SARACCO. *A new approach on recursive and non-recursive SIR methods*, in "Journal of the Korean Statistical Society", 2012, To appear [*DOI :* 10.1016/J.JKSS.2011.05.005], http://hal.inria.fr/hal-00642653/en.

[10] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Numerical method for expectations of piecewise-determistic Markov processes*, in "Communications in Applied Mathematics & Computational Science", 2012, To appear, http://hal.inria.fr/hal-00617816/en.

[11] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *Sliced Inverse Regression for stratified population*, in "Communications in Statistics - Theory and Methods", September 2011, vol. 40, n° 21, p. 3857-3878, http://hal.inria.fr/hal-00646560/en.

[12] O. COSTA, F. DUFOUR. *Singular Perturbation for the discounted continuous contol of piecewise deterministic Markov processes*, in "Applied Mathematics and Optimization", 2011, vol. 63, n° 3, p. 357-384, http://hal.inria.fr/hal-00545824/en.

[13] F. DUFOUR, T. PRIETO-RUMEAU. *Approximation of Markov Decision Processes with General State Space*, in "Journal of Mathematical Analysis and applications", 2012, To appear, http://hal.inria.fr/hal-00648223/en.

[14] F. DUFOUR, R. STOCKBRIDGE. *On the existence of strict optimal controls for constrained, controlled Markov processes in continuous-time*, in "Stochastics: An International Journal of Probability and Stochastic Processes", 2011, To appear, http://hal.inria.fr/hal-00648221/en.

[15] A. GÉGOUT-PETIT, R. AZAÏS, J. SARACCO. *Optimal quantization applied to Sliced Inverse Regression*, in "Journal of Statistical Planning and Inference", February 2012, vol. 142, n° 2, p. 481-492 [*DOI :* 10.1016/J.JSPI.2011.08.006], http://hal.inria.fr/hal-00556420/en.

[16] J. JOSSE, M. CHAVENT, B. LIQUET, F. HUSSON. *Handling missing values with Regularized Iterative Multiple Correspondence Analysis.*, in "Journal of Classification", 2011, To appear, http://hal.inria.fr/hal-00646510/en.

[17] B. YANG, G. LI, H. ZHANG, M. DAI. *Rotation and translation invariants of Gaussian-Hermite moments*, in "Pattern Recognition Letters", July 2011, vol. 32, n° 9, p. 1283-1298 [*DOI :* 10.1016/J.PATREC.2011.03.012], http://hal.inria.fr/hal-00648097/en.

[18] B. DE SAPORTA, F. DUFOUR. *Numerical method for impulse control of Piecewise Deterministic Markov Processes*, in "Automatica", 2012, to appear, http://hal.inria.fr/hal-00541413/en.

[19] B. DE SAPORTA, F. DUFOUR, H. ZHANG, C. ELEGBEDE. *Optimal stopping for predictive maintenance*, in "Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability", 2012, To appear, http://hal.inria.fr/hal-00554759/en.

[20] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*, in "Electronic Journal of Statistics", 2011, vol. 5, p. 1313-1353 [*DOI :* 10.1214/11-EJS643], http://hal.inria.fr/hal-00545447/en.

Articles in National Peer-Reviewed Journal

[21] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA, J. CLAUZEL. *Classification de données EEG par algorithme évolutionnaire pour l'étude d'états de vigilance*, in "RNTI", 2012, http://hal.inria.fr/hal-00643438/en.

Invited Conferences

[22] F. DUFOUR, B. DE SAPORTA, A. BRANDEJSKY. *Computational method for the expectation of functionals of piecewise-deterministic Markov processes*, in "Markov & semi-Markov Processes & Related Fields 2011", Greece, September 2011, http://hal.inria.fr/hal-00648224/en.

International Conferences with Proceedings

[23] R. AZAÏS, A. GÉGOUT-PETIT, J. SARACCO. *Application de la Quantification Optimale à la méthode SIR*, in "43èmes Journées de Statistique", Gammarth, Tunisia, January 2011, http://hal.inria.fr/hal-00643820/en.

[24] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR, C. ELEGBEDE. *Numerical method for the distribution of a service time*, in "ESREL 2011", France, 2011, http://hal.inria.fr/hal-00645439/en.

[25] M. PACE, H. ZHANG. *Grid based PHD filtering by Fast Fourier Transform*, in "14th International Conference on Fusion", Chicago, United States, July 2011, http://hal.inria.fr/hal-00648109/en.

[26] H. ZHANG. *Non-linear Bayesian Filtering by Convolution Method Using Fast Fourier Transform*, in "14th International Conference on Fusion", Chicago, United States, July 2011, http://hal.inria.fr/hal-00648104/en.

[27] B. DE SAPORTA, F. DUFOUR, H. ZHANG. *Approximation of the value function of an impulse control problem of Piecewise Deterministic Markov Processes*, in "18th IFAC world congress", Milano, Italy, 2011, http://hal.inria.fr/hal-00617822/en.

Conferences without Proceedings

[28] R. AZAÏS. *Estimation du taux de saut pour une classe de processus markoviens de saut*, in "Journées des jeunes statisticiens", Aussois, France, SFDS, September 2011, http://hal.inria.fr/hal-00647030/en.

[29] C. BAYSSE. *Modèle de Markov caché pour la détection d'un mode de fonctionnement dégradé d'un équipement optronique*, in "Journée des jeunes statisticiens", Aussois, France, September 2011, http://hal.inria.fr/hal-00650702/en.

[30] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Computing moments of the exit time for piecewise-deterministic Markov processes*, in "Applied Probability Society Conference", Sweden, July 2011, http://hal.inria.fr/hal-00648227/en.

[31] M. CHAVENT, S. GIRARD, V. KUENTZ, B. LIQUET, J. SARACCO. *An adaptative SIR method for block-wise evolving datastreams*, in "XIVth International Symposium of Applied Stochastic Models and Data Analysis (ASMDA 2011)", Italy, June 2011, http://hal.inria.fr/hal-00646589/en.

[32] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *Classification de variables : le package ClustOfVar*, in "43èmes Journées de Statistique (JdS)", Tunisia, May 2011, http://hal.inria.fr/hal-00646581/en.

[33] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *Clustering of variables via the PCAMIX method.*, in "Symposium of the International Federation of Classification Societies (IFCS 2011)", Frankfurt, Germany, August 2011, http://hal.inria.fr/hal-00646811/en.

[34] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *Clustering of variables via the PCAMIX method*, in "International Classication Conference", Saint Andrews, United Kingdom, July 2011, http://hal.inria.fr/hal-00646593/en.

[35] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *ClustOfVar: an R package for the clustering of variables.*, in "The R User conference", Warwick, United Kingdom, August 2011, http://hal.inria.fr/hal-00646805/en.

[36] M. CHAVENT, V. KUENTZ, J. SARACCO. *Rotation orthogonale dans PCAMIX*, in "18èmes Rencontres de la Société Francophone de Classification (SFC 2011)", Orléans, France, September 2011, http://hal.inria.fr/hal-00646814/en.

[37] R. COUDRET. *Estimateurs non paramétriques appliqués à des données valvométriques*, in "4ème Rencontre des Jeunes Statisticiens", Aussois, France, September 2011, http://hal.inria.fr/hal-00650697/en.

[38] F. DUFOUR, A. BRANDEJSKY, B. DE SAPORTA. *Numerical method for the distribution of an exit time for piecewise-deterministic Markov processes*, in "SIAM Conference on Control and Its Applications", Baltimore, United States, 2011, http://hal.inria.fr/hal-00651857/en.

[39] A. GÉGOUT-PETIT, B. DE SAPORTA, L. MARSALLE. *Bifurcating autoregressive processes with missing data and application to cell division data*, in "International Biometric Society Channel Network 3rd conference", Bordeaux, France, April 2011, http://hal.inria.fr/hal-00646334/en.

[40] A. GÉGOUT-PETIT, B. DE SAPORTA, L. MARSALLE. *Limit Theorems for bifurcating autoregressive processes with missing data and application to cell division data*, in "SIS 2011 Statistical Conference", Bologne, Italy, June 2011, http://hal.inria.fr/hal-00646337/en.

[41] J. JOSSE, M. CHAVENT, B. LIQUET, F. HUSSON. *Handling missing values with regularized iterative multiple correspondence analysis*, in "The 6th CARME conference", Rennes, France, February 2011, http://hal.inria.fr/hal-00647378/en.

[42] L. VEZARD, M. CHAVENT, P. LEGRAND, F. FAITA-AINSEBA, J. CLAUZEL. *Caractérisation d'états psychophysiologiques par classification de signaux EEG. Intégration de ces résultats dans le projet PSI.*, in "Journée AFIM : électroencéphalographie et composition musicale", Talence, France, June 2011, http://hal.inria.fr/hal-00649520/en.

[43] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Limit theorems for bifurcating autoregressive processes with missing data*, in "The Applied Probability Society Conference", Stockholm, Sweden, 2011, http://hal.inria.fr/hal-00617827/en.

Other Publications

[44] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *ClustOfVar: An R Package for the Clustering of Variables*, 2011, R Package, http://hal.inria.fr/hal-00649559/en.

[45] M. CHAVENT, V. KUENTZ, J. SARACCO. *Orthogonal rotation in PCAMIX*, 2011, R Package, http://hal.inria.fr/hal-00649561/en.

References in notes

[46] T. ALDEMIR, S. GUARRO, ET AL.. *Benchmark Implementation of Two Dynamic Methodologies for the Reliability Modeling of Digital Instrumentation and Control Systems*, NUREG/CR-6985, 2008.

[47] V. BALLY, G. PAGÈS. *A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems*, in "Bernoulli", 2003, vol. 9, n^{o} 6, p. 1003–1049.

[48] C. BAYSSE, D. BIHANIC, A. GÉGOUT-PETIT, J. SARACCO. *Modèle de Markov caché pour la détection d'un mode de fonctionnement dégradé d'un équipement optronique*, Equipe Projet Contrôle de Qualité et Fiabilité Dynamique (CQFD), INRIA Bordeaux, Thales Optronique, 2011.

[49] C. COCOZZA-THIVENT, R. EYMARD, S. MERCIER. *A finite-volume scheme for dynamic reliability models*, in "IMA J. Numer. Anal.", 2006, vol. 26, n^{o} 3, p. 446–471.

[50] M. DAVIS. *Markov models and optimization*, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1993, vol. 49.

[51] N. DUAN, K.-C. LI. *Slicing regression: a link-free regression method*, in "Ann. Statist.", 1991, vol. 19, n^{o} 2, p. 505–530, http://dx.doi.org/10.1214/aos/1176348109.

[52] R. DUDA, P. HART, D. STORK. *Pattern Classification*, John Wiley, 2001.

[53] F. DUFOUR, B. DE SAPORTA, H. ZHANG. *Controle optimal stochastique, application à l'optimisation de trajectoire, cas multicible*, Equipe Projet Contrôle de Qualité et Fiabilité Dynamique (CQFD), INRIA Bordeaux, DCNS, 2011.

[54] R. EYMARD, S. MERCIER, A. PRIGNET. *An implicit finite volume scheme for a scalar hyperbolic problem with measure data related to piecewise deterministic Markov processes*, in "J. Comput. Appl. Math.", 2008, vol. 222, n^{o} 2, p. 293–323.

[55] V. KOTHARE, C. FALINOWER. *Level Control in the Steam Generator of a Nuclear Power Plant*, in "IEEE Transactions on Control systems Technology", 2000, vol. 8, n^{o} 1.

[56] K.-C. LI. *Sliced inverse regression for dimension reduction*, in "J. Amer. Statist. Assoc.", 1991, vol. 86, n^{o} 414, p. 316–342, With discussion and a rejoinder by the author.

[57] G. PAGÈS. *A space quantization method for numerical integration*, in "J. Comput. Appl. Math.", 1998, vol. 89, n^{o} 1, p. 1–38.