IN PARTNERSHIP WITH:
CNRS

Université de Bordeaux

Activity Report 2012

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 Methods, Statistical Methods, Control Theory, Optimization, Data Analysis, Reliability

Creation of the Project-Team: January 01, 2009 .

1. Members

Faculty Members

François Dufour [Institut Polytechnique de Bordeaux, Team Leader, Professor (Pr), HdR]
Marie Chavent [University Bordeaux Segalen, Associate Professor (MdC), HdR]
Anne Gégout-Petit [University Bordeaux Segalen, Associate Professor (MdC), HdR]
Benoîte de Saporta [University Montesquieu Bordeaux IV, Associate Professor (MdC)]
Jérôme Saracco [Institut Polytechnique de Bordeaux, Professor (Pr), HdR]
Huilong Zhang [University Bordeaux 1, Associate Professor (MdC)]

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, until aug. 2012]
Isabelle Charlier [Bordeaux and Bruxelles Universités, since dec. 2012]
Raphaël Coudret [University Bordeaux 1, MENRT grant, since Oct 2010]
Amaury Labenne [IRSTEA, since sept. 2012]
Laurent Vézard [CNRS, Region grant, since Oct 2010]

Administrative Assistant

Nicolas Jahier

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 of the Year

CQFD made advances in the pratical use of its algorithms with DCNS. In the particular case of submarine command, we have coupled an tracking algorithm with an optimization code in order to compute optimal trajectories using only signals issued from embedded sonars. These results will be developped in an operating simulator.

The CQFD team created in 2012 a new annual national conference for the users of the statistical software R. The "Premières Rencontres R" are conceived as a place to present and share ideas on using the R statistical software. This meeting is designed to be a nationwide event where various topics belong, such as graphical tools, applied statistics, biostatistics, bayesian statistics, bioinformatics, data analysis, modeling, machine learning, high performance computing, etc...

The Rencontres R contained 5 guest lectures, 32 regular talks, 12 Lightning Talks and 6 posters on the following topics :

- new advances in statistics and their implementation with R,
- new R packages,
- applications or original case studies involving the R software (genetics, bioinformatics, environment, psychometrics, social sciences, neuroscience, etc...),
- computer features about the R software (multithreading, graphical tools, binding with other sofwares, etc...) ,
- topics about teaching methods with R.

This meeting was intended to everyone interested in R: researchers, teachers, people from industries, students, etc... It was built for both beginners and advanced R users, statisticians and informaticians, as well as wellwishers from every area where R can be useful. More than hundred participants attended this first edition of the conference.

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 regimes of the system and the behavior of the physical variables under each of these regimes. 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 [88] and Duan and Li [76]
  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 [77]
- 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. Dependability and safety

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.

However 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. New Results

5.1. Singularly Perturbed Discounted Markov Control Processes in a General State Space

Participant: François Dufour.

Markov decision processes, optimal control, infinite discounted expected cost, optimal control, singular perturbation

In this work, it is studied the asymptotic optimality of discrete-time Markov Decision Processes (MDP's in short) with general state space and action space and having weak and strong interactions. The idea in this work is to consider a MDP with general state and action spaces and to reduce the dimension of the state space by considering an averaged model. This formulation is often described by introducing a small parameter $\epsilon > 0$ in the definition of the transition kernel, leading to a singularly perturbed Markov model with two time scales. Our objective is twofold. First it is shown that the value function of the control problem for the perturbed system converges to the value function of a limit averaged control problem as ϵ goes to zero. In the second part of this work, it is proved that a feedback control policy for the original control problem defined by using an optimal feedback policy for the limit problem is asymptotically optimal. Our work extends existing results of the literature in the following two directions: the underlying MDP is defined on general state and action spaces and we do not impose strong conditions on the recurrence structure of the MDP such as Doeblin's condition.

These results have been obtained in collaboration with Oswaldo Luis Do Valle Costa from Escola Politécnica da Universidade de São Paulo, Brazil.

It has been published in SIAM Journal of Control and Optimization [16].

5.2. The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach.

Participant: François Dufour.

Markov decision process, expected total cost criterion, constraints, linear programming, occupation measure

This work deals with discrete-time Markov Decision Processes (MDP's) under constraints where all the objectives have the same form of an expected total cost over the infinite time horizon. The existence of an optimal control policy is discussed by using the convex analytic approach. We work under the assumptions that the state and action spaces are general Borel spaces and the model is non-negative, semi-continuous and there exists an admissible solution with finite cost for the associated linear program. It is worth noting that, in contrast with the classical results of the literature, our hypotheses do not require the MDP to be transient or absorbing. Our first result ensures the existence of an optimal solution to the linear program given by an occupation measure of the process generated by a randomized stationary policy. Moreover, it is shown that this randomized stationary policy provides an optimal solution to this Markov control problem. As a consequence,

these results imply that the set of randomized stationary policies is a sufficient set for this optimal control problem. Finally, our last main result states that all optimal solutions of the linear program coincide on a special set with an optimal occupation measure generated by a randomized stationary policy. Several examples are presented to illustrate some theoretical issues and the possible applications of the results developed in the paper.

These results have been obtained in collaboration with Alexey Piunovskiy from Department. of Mathematical Sciences, The University of Liverpool, United Kingdom and with Masayuki Horiguchi form the Department of Mathematics, Faculty of Engineering, Kanagawa University, Japan.

It has been published in Advances in Applied Probability [17] and in the invited session of the 25th conference EURO 2012 [27]

5.3. Approximation of Infinite Horizon Discounted Cost Markov Decision Processes

Participant: François Dufour.

Markov decision processes, infinite horizon discounted cost criterion, approximation and discretization

In this work, we deal with a discrete-time infinite 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.

These results have been obtained in collaboration with Tomas Prieto-Rumeau, Department of Statistics and Operations Research, UNED, Madrid, Spain.

It has been published in the book Optimization, Control, and Applications of Stochastic Systems. In Honor of Onésimo Hernandez-Lerma [52].

5.4. Continuous Control of Piecewise Deterministic Markov Processes with Long Run Average Cost

Participant: François Dufour.

Piecewise-deterministic Markov Processes, long-run average cost, optimal control, integro-differential optimality equation

The main goal of this work is to derive sufficient conditions for the existence of an optimal control strategy for the long run average continuous control problem of piecewise deterministic Markov processes (PDMP's) taking values in a general Borel space and with compact action space depending on the state variable. In order to do that we apply the so-called vanishing discount approach to obtain a solution to an average cost optimality inequality (ACOI) associated to the long run average cost problem. Our main assumptions are written in terms of some integro-differential inequalities related to the so-called expected growth condition, and geometric convergence of the post-jump location kernel associated to the PDMP.

These results have been obtained in collaboration with Oswaldo Luis Do Valle Costa from Escola Politécnica da Universidade de São Paulo, Brazil.

It has been published in the book Stochastic Processes, Finance and Control. A Festschrift in Honor of Robert J. Elliott [51].

5.5. Optimal stopping for partially observed piecewise-deterministic Markov processes

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

We have investigated an optimal stopping problem under partial observation for piecewise-deterministic Markov processes (PDMP) both from the theoretical and numerical points of view. PDMP's have been introduced by Davis [73] as a general class of stochastic models. They form a family of Markov processes involving deterministic motion punctuated by random jumps. One important property of a PDMP, relevant for the approach developed in this paper, is that its distribution is completely characterized by the embedded discrete time Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$ where Z_n is the n-th post-jump location and S_n is the n-th inter-jump time. We consider the following optimal stopping problem for a partially observed PDMP $(X_t)_{t\geq 0}$. Roughly speaking, the observation process $(Y_t)_{t\geq 0}$ is a point process defined through the embedded discrete time Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$. The inter-arrival times are given by $(S_n)_{n\in\mathbb{N}}$ and the marks by a noisy function of $(Z_n)_{n\in\mathbb{N}}$. For a given reward function g and a computation horizon $N \in \mathbb{N}$, we study the following optimal stopping problem

$$\sup_{\sigma\leq T_N} \mathbb{E}\left[g(X_\sigma)\right],$$

where T_N is the N-th jump time of the PDMP $(X_t)_{t\geq 0}$, σ is a stopping time with respect to the natural filtration $\mathcal{F}^o = (\mathcal{F}^o_t)_{t\geq 0}$ generated by the observations $(Y_t)_{t\geq 0}$.

A general methodology to solve such a problem is to split it into two sub-problems. The first one consists in deriving the filter process given by the conditional expectation of X_t with respect to the observed information $\mathcal{F}^o_t$. Its main objective is to transform the initial problem into a completely observed optimal stopping problem where the new state variable is the filter process. The second step consists in solving this reformulated problem, the new difficulty being its infinite dimension. Indeed, the filter process takes values in a set of probability measures.

Our work is inspired by [92] which deals with an optimal stopping problem under partial observation for a Markov chain with finite state space. The authors study the optimal filtering and convert their original problem into a standard optimal stopping problem for a continuous state space Markov chain. Then they propose a discretization method based on a quantization technique to approximate the value function. However, their method cannot be directly applied to our problem for the following main reasons related to the specificities of PDMPs.

Firstly, PDMPs are continuous time processes. Then, it appears natural to work with the embedded Markov chain $(Z_n, S_n)_{n\in\mathbb{N}}$. In addition, we assume that $(Z_n)_{n\in\mathbb{N}}$ takes finitely many values. However, an important difficulty is that the structure of stopping time remains intrinsically continuous. Consequently, our problem cannot be converted into a fully discrete time problem.

Secondly, the distribution of a PDMP combines both absolutely continuous and singular components. This is due to the existence of forced jumps when the process hits the boundary of the state space. As a consequence the derivation of the filter process is not straightforward. In particular, the absolute continuity hypothesis **(H)** of [92] does not hold.

Thirdly, in our context the reformulated optimization problem is not standard, unlike in [92]. Indeed, although we obtain a reformulation similar to an optimal stopping problem for a fully observed PDMP, it involves the Markov chain $(\Pi_n, S_n)_{n\in\mathbb{N}}$ that is not the embedded Markov chain of some PDMP. Therefore, a new derivation of dynamic programming equations is required as we cannot use the results of [81]. In particular, one needs to derive fine properties of the structure of the $(\mathcal{F}^o_t)_{t\geq 0}$-stopping times. Moreover, we construct an ϵ-optimal stopping time.

Finally, a natural way to proceed with the numerical approximation is then to follow the ideas developed in [92] [8] namely to replace the filter Π_n and the inter-jump time S_n by some finite state space approximations in the dynamic programming equation. However, a noticeable difference from [8] lies in the fact that the dynamic programming operators therein were Lipschitz continuous whereas our new operators are only Lipschitz continuous between some points of discontinuity. We overcome this drawback by splitting the operators into their restrictions onto their continuity sets. This way, we obtain not only an approximation of the value function

of the optimal stopping problem but also an ϵ-optimal stopping time with respect to the filtration $(\mathcal{F}_t^o)_{t\geq 0}$ that can be computed in practice.

This work is submitted for publication [60] and presented in an invited international conference [26].

5.6. Predictive maintenance for the heated hold-up tank

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

A complex system is inherently sensitive to failures of its components. One must therefore determine maintenance policies in order to maintain an acceptable operating condition. Optimizing the maintenance is a very important problem in the analysis of complex systems. It determines when it is best that maintenance tasks should be performed on the system in order to optimize a cost function: either maximize a performance function or conversely minimize a loss function. Moreover, this optimization must take into account the random nature of failures and random evolution and dynamics of the system.

The example considered here is the maintenance of the heated hold-up tank, a well know test case for dynamic reliability, see e.g. [75], [89], [90], [94]. The system consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps and one outlet valve. A thermal power source heats up the fluid. The failure rate of the components depends on the temperature, the position of the three components monitors the liquid level in the tank, and in turn, the liquid level determines the temperature. The main characteristic of this system is that it can be modeled by a stochastic hybrid process, where the discrete and continuous parts interact in a closed loop. As a consequence, simulating this process and computing related reliability indices has been a challenge for the dynamic reliability community. To our best knowledge, optimization of maintenance policies for the heated hold-up tank has not been addressed yet in the literature.

The only maintenance operation considered here is the complete replacement of all the failed components and the system restarts in its initial equilibrium state. Partial repairs are not allowed. Mathematically, this problem of preventive maintenance corresponds to a stochastic optimal stopping problem as explained by example in the book of Aven and Jensen [68]. It is a difficult problem because of the closed loop interactions between the state of the components and the liquid level and temperature. A classical approach consists in using condition-based maintenance (CBM) to act on the system based on its current state and before its failure. One can for example calculate the remaining useful life (RUL) of the system and the preventive replacement is carried out when the deterioration level exceeds a certain threshold or enters in a certain state [96], [80]. Our approach also takes into account the current state of the process, but our decision rule is not based on damage accumulation nor does it correspond to hitting some threshold. Instead, it involves a performance function that reflects that the longer the system is in a functioning state the better.

The dynamics of the heated hold-up tank can be modeled by a piecewise deterministic Markov process (PDMP), see [94]. Therefore, our maintenance problem boils down to an optimal stopping problem for PDMP's. PDMP's are a class of stochastic hybrid processes that has been introduced by Davis [73] in the 80's. These processes have two components: a Euclidean component that represents the physical system (e.g. temperature, pressure, ...) and a discrete component that describes its regime of operation and/or its environment. Starting from a state x and mode m at the initial time, the process follows a deterministic trajectory given by the laws of physics until a jump time that can be either random (e.g. it corresponds to a component failure or a change of environment) or deterministic (when a magnitude reaches a certain physical threshold, for example the pressure reaches a critical value that triggers a valve). The process restarts from a new state and a new mode of operation, and so on. This defines a Markov process. Such processes can naturally take into account the dynamic and uncertain aspects of the evolution of the system. A subclass of these processes has been introduced by Devooght [75] for an application in the nuclear field. The general model has been introduced in dynamic reliability by Dutuit and Dufour [79].

As illustrated above, it is crucial to have an efficient numerical tool to compute the optimal maintenance time in practical cases. To this aim, a general numerical approach was developed in [8]. It was first applied to an example of maintenance of a metallic structure subject to corrosion, without closed loop interactions or deterministic jumps, and with a simple cost function that did not depend on time, see [23]. The objective of

the present paper is to further demonstrate the high practical power of the theoretical methodology described in [8], by applying it to the more challenging heated hold-up tank problem. The cost function chosen here is also more complex as it takes into account both continuous components as well as the running time. More precisely, we propose to compute the optimal cost as well as a quasi-optimal stopping rule, which is the date when the maintenance should be performed. As a by-product of our procedure, the distribution of the optimal maintenance dates is also obtained, as well as the distributions of the liquid level and temperature at the chosen maintenance date.

This work is submitted for publication [66] and presented in an international conference [32].

5.7. 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). We have proposed 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 [67] 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 [87]. 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 [73] 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. In 2012, we have further improved our simulator by taking captors (and their possible failures) into account.

This work was presented in an international conference [36], a national conference [39] and is published as a book chapter [49].

5.8. Stochastic control for underwater optimal trajectories

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

This work aims to compute optimal trajectories for underwater vehicles evolving in a given environment to accomplish some tasks. This is an optimal control problem. In real context, available inputs are not perfectly known. Hence a stochastic approach seems to be needed. Markov decision processes (MDPs) constitute a general family of controlled stochastic processes suitable for the modeling of sequential decision-making problems. The analysis of MDPs leads to mathematical and computational problems. The corresponding theory has reached a rather high degree of maturity, although the classical tools (such as value iteration, policy iteration, linear programming, and their various extensions) are generally hardly applicable in practice. Hence, solving MDPs numerically is an awkward and important problem. The method is applied to control a submarine which wants to well detect one or several targets. Why? A smart operator, if provided information about target's position and velocity and a sound propagation code can find a good trajectory. If we- now consider a submarine surrounded by several targets, it is clear that a human operator will have great difficulty to find the best route.

This work was presented in an international conference [35].

5.9. Statistical study od asymmetry in cell lineage data

Participants: Benoîte de Saporta, Anne Gégout-Petit.

This work proposes a rigorous methodology to study cell division data consisting in several observed genealogical trees of possibly different shapes. For instance, [93] filmed 94 colonies of Escherichia coli cells dividing between four and nine times. We propose a new rigorous approach to take into account all the available information. Indeed, we propose an inference based on a finite fixed number of replicated trees when the total number of observed cells tends to infinity. We use the missing data asymmetric BAR model introduced by [7]. In this approach, the observed genealogies are modeled with a two-type Galton Watson (GW) process. However, we propose a different least-squares estimator for the parameters of the BAR process that does not correspond to the single-tree estimators averaged on the replicated trees. We also propose an estimator of the parameters of the GW process specific to our binary tree structure and not based simply on the observation of the number of cells of each type in each generation.

Our procedure allows us to fully take into account missing observations, data from different trees as well as the dependence structure within genealogical trees. It also enables us to use all the information available without the drawbacks of low accuracy for estimators or low power for tests on small single trees. We study the consistency and asymptotic normality of our estimators and derive asymptotic confidence intervals as well as Wald's type tests to investigate the asymmetry of the data for both the BAR and GW processes. Our results are applied to the Escherichia coli data of [93].

This work is in collaboration with Laurence Marsalle (Lille 1 University). It is submitted for publication [65] and was presented in an international conference [33].

5.10. Random coefficient bifurcating autoregressive processes

Participants: Benoîte de Saporta, Anne Gégout-Petit.

In the 80's, Cowan and Staudte [72] introduced Bifurcating Autoregressive processes (BAR) as a parametric model to study cell lineage data. A quantitative characteristic of the cells (e.g. growth rate, age at division) is recorded over several generations descended from an initial cell, keeping track of the genealogy to study inherited effects. As a cell usually gives birth to two offspring by division, such genealogies are naturally structured as binary trees. BAR processes are thus a generalization of autoregressive processes (AR) to this binary tree structure, by modeling each line of descent as a first order AR process, allowing the environmental effects on sister cells to be correlated. Statistical inference for the parameters of BAR processes has been widely studied, either based on the observation of a single tree growing to infinity [72], [85], [83], [95] or on a large number of small independent trees [86], [84].

Various extensions of the original model have been proposed, but to our best knowledge, only two papers [71] and [70] deal with random coefficient BAR processes. In the former by Bui and Huggins it is explained that random coefficients BAR processes can account for observations that do not fit the usual BAR model. For instance, the extra randomness can model irregularities in nutrient concentrations in the media in which the cells are grown. In this work, we propose a new model for random coefficient BAR processes (R-BAR). It is more general than that of Bui and Huggins, as the random variables are not supposed to be Gaussian, they may not have moments of all order and correlation between all the sources of randomness are allowed. Moreover, we propose an asymmetric model in the continuance of [82], [69], [74], [70], [7], [24] in the context of missing data. Indeed, experimental data are often incomplete and it is important to take this phenomenon into account for the inference. We model the structure of available data by a Galton Watson tree, instead of a complete binary tree. Our model is close to that developed in [70], but the assumptions on the noise process are different as we allow correlation between the two sources of randomness but require higher moments because of the missing data and because we do not use a weighted estimator. The main difference is that the model in [70] is fully observed, whereas ours allows for missing observations.

Our approach for the inference of our model is also different from [71], [70]. As we cannot use maximum likelihood estimation, we propose modified least squares estimators as in [91]. The originality of our approach is that it combines the bifurcating Markov chain and martingale approaches. Bifurcating Markov chains (BMC) were introduced in [82] on complete binary trees and further developed in [74] in the context of missing data on Galton Watson trees. BAR models can be seen as a special case of BMC. This interpretation allows us to establish the convergence of our estimators. A by-product of our procedure is a new general result for BMC on Galton Watson trees. Indeed, in [82], [74] the driven noise sequence is assumed to have moments of all order. Here, we establish new laws of large numbers for polynomial functions of the BMC where the noise sequence only has moments up to a given order. The strong law of large numbers [78] and the central limit theorem for martingales have been previously used in the context of BAR processes and adapted to special cases of martingales on binary trees. In this paper, we establish a general law of large numbers for square integrable martingales on Galton Watson binary trees. This result is applied to our R-BAR model to obtain sharp convergence rates and a quadratic strong law for our estimators.

This work is in collaboration with Laurence Marsalle (Lille 1 University). It is submitted for publication [64].

5.11. Hidden Markov Model for the detection of a degraded state in an optronic equipment

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

As part of optimizing the reliability, Thales Optronics now includes systems that examine the state of its equipment. This function is performed by HUMS (Health & Usage Monitoring System). We hope to implement a program based on these observations that can determine the lifetime of this optronic equipment. 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 system. Using this information about this variable, we seek to detect as soon as possible a degraded state and propose maintenance before failure. For this we use a hidden Markov model. The state of our system at time t is then modeled by a Markov chain X_t. 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 breaking down at time t, knowing the history of the TMF until this moment. We have subsequently studied this methodology with simulated data. Then finally we applied these results on the analysis of our real data and we have checked that the results are consistent with the reality. So using this method could allow the company to recall equipments which are estimated in deteriorated state and do not control those estimated in stable state. Thales Optronics could improve its maintenance system and reduce its cost function.

This work is a part of the CIFRE PhD of Camille Baysse also supervised for the Thales part by Didier Bihannic and Michel Prenat. It was presented in an national conference [38] and is submitted for publication in an national per-reviewed journal [58].

5.12. Predictive maintenance for an optronic equipment

Participants: Camille Baysse, Benoîte de Saporta, Anne Gégout-Petit, Jérôme Sarraco.

After the problem of detection of a degraded state, we have tackled the problem of predictive maintenance for an optronic equipment. For this we model the state of the system by a PDMP (state with three possible values and cumulative time of use). In this framework, we reformulate the problem of maintenance of optimization in an optimal stopping problem maximizing a criteria about time of use without failure. In this framework, we can use known results developed in the CQFD team on optimal control [8], [23]. We have extensively studied the problem with simulated data, computed grid of quantization and optimal policy for the real problem. This results will be implemented by Thales in HUMS of optronic equipment.

This work was presented in an national conference [38] and an abstract is accepted for publication in an international conference with papers.

5.13. Non parametric estimation of the jump rate for non-homogeneous marked renewal processes

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

This work is devoted to the nonparametric estimation of the jump rate and the cumulative rate for a general class of non-homogeneous marked renewal processes, defined on a separable metric space. In our framework, the estimation needs only one observation of the process within a long time. Our approach is based on a generalization of the multiplicative intensity model, introduced by Aalen in the seventies. We provide consistent estimators of these two functions, under some assumptions related to the ergodicity of an embedded chain and the characteristics of the process. The methodology is illustrated by a numerical example. It is the object of a paper [57] to appear in the Annales de l'Intitut Poincaré

5.14. Non parametric estimation of conditional distribution of the interjumping times for piecewise Markov processes

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

This work gives a nonparametric method for estimating the conditional density associated to the jump rate of a piecewise-deterministic Markov process. In our framework, the estimation needs only one observation of the process within a long time interval. Our method relies on a generalization of Aalen's multiplicative intensity model. We prove the uniform consistency of our estimator, under some reasonable assumptions related to the primitive characteristics of the process. A simulation example illustrates the behavior of our estimator. This work is the object of a paper [56] submitted for publication

5.15. Stochastic modelling and simulation of fatigue crack propagation using piecewise-deterministic Markov processes

Participants: Romain Azaïs, Anne Gégout-Petit.

Fatigue crack propagation is a stochastic phenomenon in nature due to the inherent uncertainties coming from material properties, environmental conditions and loads. Stochastic processes offer an appropriate framework for modelling crack propagation since it is intended to include sources variabilities. In this work, we propose to model crack propagation mechanism with Piecewise Deterministic Markov Process (PDMP) using usual random crack laws. Conventional laws proposed in the literature seem inadequate for describing the whole fatigue crack trajectory mainly when the crack extends in a rapid manner. To overcome this drawback, a new modelling is proposed that consists in using more than one law as each one is more suitable for a specific phase during crack propagation. Regime-switching models seem very attractive and with our modelling assessed crack growth rates and crack lengths are very close to experimental values. Moreover, behaviour just before failure is well captured and can be discussed. Empirical curves from literature are used to adjust the parameters associated to the proposed modelling. Statistical observations and numerical simulations show the efficiency of the proposed approach to model and to simulate fatigue crack growth. This work has been presented in an international congress [34] and is the object of a paper which will be submitted very soon.

5.16. Statistical Analysis of Grapevine Mortality Associated with Esca or Eutypa Dieback Foliar Expression

Participant: Anne Gégout-Petit.

Esca and Eutypa dieback are two major wood diseases of grapevine in France. Their widespread distribution in vineyards leads to vine decline and to a loss in productivity. However, little is known either about the temporal dynamics of these diseases at plant level, and equally, the relationships between foliar expression of the diseases and vine death is relatively unknown too. . To investigate these questions, we surveyed the vines of six vineyards cv. Cabernet Sauvignon in the Bordeaux region, by recording foliar symptoms, dead arms and dead plants from 2004 to 2010. In 2008, 2009 and 2010, approximately five percent of the asymptomatic vines died but the percentage of dead vines which had previously expressed esca foliar symptoms was higher, and varied between vineyards. A logistic regression model was used to select the previous years of symptomatic expression associated with vine mortality. The mortality of esca is always associated with the foliar symptom expression of the year preceding vine death. One or two other earlier years of expression frequently represented additional risk factors. The Eutypa dieback symptom was also a risk factor of death, superior or equal to that of esca. The study of the internal necroses of vines expressing esca or Eutypa dieback is discussed in the light of these statistical results. This work has been presented in an international congress [44] and is the object of a submitted paper.

5.17. MonteCarlo test for two patterns of point processes on a grid

Participants: Anne Gégout-Petit, Marie Chavent, Amaury Labenne.

In order to compare two patterns of distribution of symptomatic or dead vines in a same vineyard but for two consecutive years, we have developed a Monte Carlo test. First we estimate the intensity of occurrence of disease in one of the pattern, then we simulate n realizations i.i.d. of this intensity and compute the associate likelihoods in order to build an interval that cover $(1-\alpha)$ per cent of the realizations. The test reject the equality of repartition if the likelihood computed with the second pattern is not included in this interval. We have made simulations and applied this test to the repartition of esca in vineyard. This work has been presented in a national workshop on software R [46].

5.18. 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 it seems 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 standard multivariate analysis techniques, and classification, we have proposed criteria to discriminate groups of patients different in regard of their response to treatment. 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 submitted paper in a international peer-reviewed journal in the domain of pneumology .

5.19. A hidden renewal model for monitoring aquatic systems biosensors

Participants: Romain Azaïs, Raphaël Coudret.

This work aims at modeling signals of oysters' openings over time using a four-state renewal process. Two of them are of particular interest and correspond to instants when the animals are open or closed. An estimator of the cumulative jump rate of this process is provided. It relies on observations of the jumps between the four states. Here these measures are not available but the observed signal takes ranges of real values according to this underlying process. A procedure to estimate a probability density function that summarizes the information of the signal is explained. This leads to estimate the hidden renewal process and then its cumulative jump rate for each oyster. A classification of these functions for a group of oysters discriminate them according to their assumed health status. Such a diagnosis is essential when using these animals as biosensors for water quality assessment. This work is a joint work with Gilles Durrieu from Université de Bretagne Sud and in collaboration with UMR CNRS 5805 EPOC.

5.20. A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process

Participant: Romain Azaïs.

We investigate a nonparametric approach to provide a recursive estimator of the transition density of a non-stationary piecewise-deterministic Markov process, from only one observation of the path within a long time. In this framework, we do not observe a Markov chain with transition kernel of interest. Fortunately, one may write the transition density of interest as the ratio of the invariant distributions of two embedded chains of the process. Our method consists in estimating these invariant measures. We state a result of consistency under some general assumptions about the main features of the process. A simulation study illustrates the well asymptotic behavior of our estimator. This work is the object of a paper [55] submitted for publication.

5.21. A new sliced inverse regression method for multivariate response

Participants: Jérôme Saracco, Raphaël Coudret.

We consider a semiparametric regression model of a q-dimensional multivariate response y on a p-dimensional covariate x. In this paper, a new approach is proposed based on sliced inverse regression for estimating the e ffective dimension reduction (EDR) space without requiring a prespeci ed parametric model. The convergence at rate square root of n of the estimated EDR space is shown. We discuss the choice of the dimension of the EDR space. The numerical performance of the proposed multivariate SIR method is illustrated on a simulation study. Moreover, we provide a way to cluster components of y related to the same EDR space. One can thus apply properly multivariate SIR on each cluster instead of blindly applying multivariate SIR on all components of y. An application to hyperspectral data is provided.

These results have been obtained in collaboration with Stéphane Girard (Inria Rhône Alpes).

The paper is under revision for possible publication in CSDA [63].

5.22. Comparison of kernel density estimators with assumption on number of modes

Participants: Jérôme Saracco, Raphaël Coudret.

A data-driven bandwidth choice for a kernel density estimator called critical bandwidth is investigated. This procedure allows the estimation to have as many modes as assumed for the density to estimate. Both Gaussian and uniform kernels are considered. For the Gaussian kernel, asymptotic results are given. For the uniform kernel, an argument against these properties is mentioned. These theoretical results are illustrated with a simulation study which compare the kernel estimators that rely on critical bandwidth with another one which uses a plug-in method to select its bandwidth. An estimator that consists in estimates of density contour clusters and takes assumptions on number of modes into account is also considered. Finally, the methodology is illustrated using environment monitoring data.

These results have been obtained in collaboration with Gilles Durrieu (Université Bretagne-Sud).

The paper is under revision for possible publication in Communications in Statistics - Simulation and Computation [62].

5.23. A new approach on recursive and non-recursive SIR methods

Participant: Jérôme Saracco.

We consider a semiparametric single index regression model involving a p-dimensional quantitative covariable x and a real dependent variable y. A dimension reduction is included in this model via an index $x'\beta$. Sliced inverse regression (SIR) is a well-known method to estimate the direction of the Euclidean parameter β which is based on a "slicing step" of y in the population and sample versions. The goal of this paper is twofold. On the one hand, we focus on a recursive version of SIR which is also suitable for multiple indices model. On the other hand, we propose a new method called SIRoneslice when the regression model is a single index model. The SIRoneslice estimator of the direction of β is based on the use of only one "optimal" slice chosen among the H slices. Then, we provide its recursive version.We give an asymptotic result for the SIRoneslice approach. Simulation study shows good numerical performances of the SIRoneslice method and clearly exhibits the main advantage of using recursive versions of the SIR and SIRoneslice methods from a computational time point of view. A real dataset is also used to illustrate the approach. Some extensions are discussed in concluding remarks. The proposed methods and criterion have been implemented in R and the corresponding codes are available from the authors.

These results have been obtained in collaboration with Bernad Bercu (Université Bordeaux 1) and Thi Mong Ngoc Nguyen (Université de Strasbourg).

The paper has been published in the Journal of the Korean Statistical Society [11].

5.24. On the asymptotic behavior of the Nadaraya-Watson estimator associated with the recursive SIR method

Participant: Jérôme Saracco.

We investigate the asymptotic behavior of the Nadaraya-Watson estimator for the estimation of the regression function in a semiparametric regression model. On the one hand, we make use of the recursive version of the sliced inverse regression method for the estimation of the unknown parameter of the model. On the other hand, we implement a recursive Nadaraya-Watson procedure for the estimation of the regression function which takes into account the previous estimation of the parameter of the semiparametric regression model.We establish the almost sure convergence as well as the asymptotic normality for our Nadaraya-Watson estimator. We also illustrate our semiparametric estimation procedure on simulated data.

These results have been obtained in collaboration with Bernad Bercu (Université Bordeaux 1) and Thi Mong Ngoc Nguyen (Université de Strasbourg) .

The paper is submitted [59].

5.25. Comparison of sliced inverse regression approaches for underdetermined cases

Participants: Jérôme Saracco, Raphaël Coudret.

Among methods to analyze high-dimensional data, the sliced inverse regression (SIR) is of particular interest for non-linear relations between the dependent variable and some indices of the covariate. When the dimension of the covariate is greater than the number of observations, classical versions of SIR cannot be applied. Various upgrades were then proposed to tackle this issue such as RSIR and SR-SIR, to estimate the parameters of the underlying model and to select variables of interest. In this paper, we introduce two new estimation methods respectively based on the QZ algorithm and on the Moore-Penrose pseudo-inverse. We also describe a new selection procedure of the most relevant components of the covariate that relies on a proximity criterion between submodels and the initial one. These approaches are compared with RSIR and SR-SIR in a simulation study. Finally we applied SIR-QZ and the associated selection procedure to a genetic dataset in order to find eQTL.

These results have been obtained in collaboration with Benoit Liquet (Université Bordeaux 2) . The paper is submitted .

5.26. Orthogonal rotation in PCAMIX

Participants: Marie Chavent, Jérôme Saracco.

Kiers (1991) considered the orthogonal rotation in PCAMIX, a principal component method for a mixture of qualitative and quantitative variables. PCAMIX includes the ordinary Principal Component Analysis (PCA) and Multiple Correspondence Analysis (MCA) as special cases. In this work, 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".

These results have been obtained in collaboration with Vanessa Kuentz of IRSTEA (UR ABDX).

It has been published in Advances in Data Analysis and Classification [15] and presented in the context of application in cultural sociology in the Premières Rencontres R [42].

5.27. A sliced inverse regression approach for data stream

Participants: Marie Chavent, Jérôme Saracco.

In this work, we focus on data arriving sequentially by block in a stream. A semiparametric regression model involving a common EDR (Effective Dimension Reduction) direction is assumed in each block. Our goal is to estimate this direction at each arrival of a new block. A simple direct approach consists in pooling all the observed blocks and estimate the EDR direction by the SIR (Sliced Inverse Regression) method. But some disadvantages appear in practice such as the storage of the blocks and the running time for high dimensional data. To overcome these drawbacks, we propose an adaptive SIR estimator of based on the SIR approach for a stratified population developed by Chavent et al. (2011). The proposed approach is faster both from computational complexity and running time points of view, and provides data storage benefits. We show the consistency of our estimator at the root-n rate and give its asymptotic distribution. We propose an extension to multiple indices model. We also provide a graphical tool in order to detect if a drift occurs in the EDR direction or if some aberrant blocks appear in the data stream. In a simulation study, we illustrate the good numerical behavior of our estimator. One important advantage of this approach is its adaptability to changes in the underlying model. Finally we apply it on real data concerning the estimation of Mars surface physical properties.

This work is under revision in Statistics and Computing [61].

5.28. ClustOfVar: An R Package for the Clustering of Variables

Participants: Marie Chavent, Jérôme Saracco.

Clustering of variables is as a way to arrange variables into homogeneous clusters, i.e., groups of variables which are strongly related to each other and thus bring the same information. These approaches can then be useful for dimension reduction and variable selection. Several specific methods have been developed for the clustering of numerical variables. However concerning qualitative variables or mixtures of quantitative and qualitative variables, far fewer methods have been proposed. The R package ClustOfVar was specifically developed for this purpose. The homogeneity criterion of a cluster is defined as the sum of correlation ratios (for qualitative variables) and squared correlations (for quantitative variables) to a synthetic quantitative variable, summarizing "as good as possible" the variables in the cluster. This synthetic variable is the first principal component obtained with the PCAMIX method. Two clustering algorithms are proposed to optimize the homogeneity criterion: iterative relocation algorithm and ascendant hierarchical clustering. We also propose a bootstrap approach in order to determine suitable numbers of clusters. We illustrate the methodologies and the associated package on small datasets.

These results have been obtained in collaboration with Vanessa Kuentz of IRSTEA (UR ABDX).

It has been published in Journal of Statistical Softwares [14]. The study of the inclusion of environment by the farmer with ClustOfVar has been presented in the Premières Rencontres R and in [45]

5.29. Divisive Monothetic Clustering for Interval and Histogram-valued Data

Participant: Marie Chavent.

In this paper we propose a divisive top-down clustering method designed for interval and histogram-valued data. The method provides a hierarchy on a set of objects together with a monothetic characterization of each formed cluster. At each step, a cluster is split so as to minimize intra-cluster dispersion, which is measured using a distance suitable for the considered variable types. The criterion is minimized across the bipartitions induced by a set of binary questions. Since interval-valued variables may be considered a special case of histogram-valued variables, the method applies to data described by either kind of variables, or by variables of both types. An example illustrates the proposed approach.

These results have been obtained in collaboration with Paula Brito of Porto University and presented in ICPRAM'2012 [31].

5.30. Classification of EEG signals by an evolutionary algorithm

Participants: Marie Chavent, Laurent Vézard.

The goal is to predict the alertness of an individual by analyzing the brain activity through electroencephalographic data (EEG) captured with 58 electrodes. Alertness is characterized as a binary variable that can be in a normal or relaxed state. We collected data from 44 subjects before and after a relaxation practice, giving a total of 88 records. After a pre-processing step and data validation, we analyzed each record and discriminate the alertness states using our proposed slope criterion. Afterwards, several common methods for supervised classification (k nearest neighbors, decision trees -CART-, random forests, PLS and discriminant sparse PLS) were applied as predictors for the state of alertness of each subject. The proposed slope criterion was further refined using a genetic algorithm to select the most important EEG electrodes in terms of classification accuracy. Results shown that the proposed strategy derives accurate predictive models of alertness.

These results have been obtained in collaboration with Pierrick Legrand of ALEA Inria team.

It has been published in Journal des Nouvelles Technologies [25] and presented in COMPSTAT 2012 [47].

5.31. Variable selection by genetic algorithm for the study of alertness states.

Participants: Marie Chavent, Laurent Vézard.

The aim of this work is to predict the state of alertness of an individual (binary variable, "normal" or "relaxed") from the study of brain activity (electroencephalographic signals EEG) collected with a limited number of electrodes. In fact, the set up of electrodes during the EEG signal acquisition is time consuming and these electrodes are correlated. In our study, the EEG of 58 participants in the two alertness states (116 records) were collected via a cap with 58 electrodes. After a data validation step based on the study of the contingent negative variation (CNV), 19 subjects were retained in the study. A CSP (Common Spacial Pattern) coupled to a linear discriminant analysis were used to build a decision rule and thus predict the alertness of the participants. A genetic algorithm was used to determine a subset of electrodes of size p '(where p' <p, where p = 58 is the number of electrodes). This presentation will present the CSP in the general framework and will introduce innovations made to this method. The genetic algorithm will be described proposed and recent results will be presented.

These results have been obtained in collaboration with Pierrick Legrand of ALEA Inria team.

It has been presented in the Journée Évolutionnaire Thématique, 23éme édition [48].

5.32. Handling Missing Values with Regularized Iterative Multiple Correspondence Analysis

Participant: Marie Chavent.

A common approach to deal with missing values in multivariate exploratory data analysis consists in minimizing the loss function over all non-missing elements. This can be achieved by EM-type algorithms where an iterative imputation of the missing values is performed during the estimation of the axes and components. This paper proposes such an algorithm, named iterative multiple correspondence analysis, to handle missing values in multiple correspondence analysis (MCA). This algorithm, based on an iterative PCA algorithm, is described and its properties are studied. We point out the over tting problem and propose a regularized version of the algorithm to overcome this major issue. Finally, performances of the regularized iterative MCA algorithm (implemented in the R-package named missMDA) are assessed from both simulations and a real dataset. Results are promising with respect to other methods such as the missing-data passive modi ed margin method, an adaptation of the missing passive method used in Gini's Homogeneity analysis framework.

It has been published in Journal of Classification [21].

6. Bilateral Contracts and Grants with Industry

6.1. 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 funded the PhD Thesis of Adrien Brandejsky (2009-2012) and is a partner of ANR Fautocoes.

6.2. 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 2012, we extended our previous results to multiple target vessels and 3D control. We also coupled our code with the output of a tracking software to take more realistically into account the uncertainty on the position and speed of the targets. This work gave rise to a new technical report [54] and a presentation in an international conference [35].

6.3. 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. We also take into account captors and their possible failures. This work gave rise to a technical report [53] and was presented in an international conference [36], a national conference [39] and is published as a book chapter [49].

6.4. Thales Optronique

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

Integrated maintenance, failure intensity, optimisation.

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 problem. This work gave rise to a technical report, was presented in an international conference [30], a national conference [38] and is submitted for publication

6.5. LyRE

Participant: Jérôme Saracco.

The goal of this contract with the LyRE ((R & D research center of Lyonnaise des Eaux)) is to provide management and consultancy tools to keep the immense heritage of these drinking water and sanitation distribution networks in optimal condition. A PhD student (K. Claudio) of J. Saracco is working in the LyRE team on sampling problems coming from a partial automatic teletransmission of water consumption data. This works has been presented at the “7ème colloque francophone sur les sondages” on November 2012 at ENSAI, Rennes.The lecture is untilted “Estimation de la consommation d’eau d’un secteur à partir d’un échantillon d’usagers télérelevés”.

A patent describing the statistical methodology has also been registered in Novembrer 2012.

7. Partnerships and Cooperations

7.1. Regional Initiatives

In collaboration with UMR SAVE of INRA de Bordeaux, Anne Gégout-Petit and Marie Chavent supervise a PhD until september 2012 founded by a regional grant on the subject "Détermination des facteurs environnementaux et culturaux liés à l’esca de la vigne par une approche de modélisation spatio temporelle".
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*.

7.2. National Initiatives

7.2.1. 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/.

7.2.2. 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.

7.3. International Initiatives

7.3.1. Collaborations with Major European Organizations

Numerical methods for Markov decision processes This research project is concerned with *numerical methods for Markov decision processes* (MDPs). Namely, we are interested in *approximating numerically* the optimal value function and the optimal controls for different classes of *constrained* and *unconstrained* MDPs. Our methods are based on combining the *linear programming (LP) formulation* of an MDP with a discretization procedure —referred to as *quantization*— of a probability distribution, underlying the random transitions of the dynamic system. We are concerned with optimality criteria such as the total expected cost criterion (for finite horizon problems) and, on the other hand, the total expected discounted cost and the average cost optimality criteria (for infinite horizon problems).

This project is supported by the *Gobierno de Espana, Derccion Genral de Investigacion Cinetifica y Tecnica* (reference number: MTM2012-31393) for three years (25 000 euros) to support the scientific collaboration between Tomas Prieto-Rumeau and François Dufour.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

Tomas Prieto-Rumeau (Department of Statistics and Operations Research, UNED, Madrid, Spain) visited the team during one month in 2012. The main subject of the collaboration is the approximation of Markov Decision Processes.

Oswaldo Costa (Escola Politécnica da Universidade de São Paulo, Brazil) collaborate with the team on the theoretical aspects of Markov Decision Processes. He visited the team during two weeks in 2012.

Alexey Piunovskiy (University of Liverpool) visited the team during one month in 2012. The main subject of the collaboration is the linear programming approach for Markov Decision Processes.

7.4.2. Visits to International Teams

François Dufour has visited A. Piunovskiy at Liverpool University for a week in March.

Jérôme Sracco was invited to MCR Biostatistics Units at Cambridge University for one week in november 2012. He gave a seminar untitled "Dimension reduction based on sliced inverse regression (SIR): a look at the special case when $n < p$.

8. Dissemination

8.1. Editorial activities

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.

8.2. Scientific responsibilities

F. Dufour is the leader of the ANR project FAUTOCOES. B. de Saporta is in charge of the tâche 3 of the ANR project FAUTOCOES.

8.3. Organization of workshops and conferences

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

8.4. Administration of the universities and research institutes

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

F. Dufour is member of the scientific council of the Institute of Mathematics of Bordeaux.

F. Dufour is vice-president of the Inria Project Committee.

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

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 correspondent of the cursus *Ingénierie Economique* of the master MIMSE *Ingénierie Mathématique, Statistique et Economique* of the University of Bordeaux.

A. Gégout-Petit is elected member of the CEVU of University Bordeaux Segalen

A. Gégout-Petit is in charge to promote diplomas of UFR Science et Modélisation.

A. Gégout-Petit is member of the Mathematical Institute of Bordeaux council

A. Gégout-Petit is general secretary and elected member of the council of the Société Française de Statistique.

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.

J. Saracco is member of the commission Inria ”Jeunes Chercheurs”.

J. Saracco is member of the council of ENSC

J. Saracco is the leader of the team ”Statistics and Probability” of the Institute of Mathematics of Bordeaux (IMB).

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

8.5. Scientific Animation

J. Saracco was member of the AERES visiting committee of the research unit of the "Laboratoire de Probabilités et modèles aléatoires" (LPMA Universités Paris 6 et Paris 7)

B de Saporta, J. Saracco and M. Chavent are elected (deputy) member of the CNU 26.

B. de Saporta belongs to the board of SMAI-MAS group.

A. Gégout-Petit was in the organizing committee of the first "Forum Emploi Mathématique" which was successful and had gather together a thousand of participants in january.

A. Gégout-Petit made a action in order to promote scientific studies with a class of secondary (statistical project with the pupils and presentation of a research project for the pupils in the university.

L. Vézard has presented the objectives of his PhD work and its contribution in the global project of musical informatics to a class of secondary school.

8.6. Teaching - Supervision - Juries

8.6.1. Teaching

Licence : F. Dufour, 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.

Licence : F. Dufour, 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.

Licence : A. Gégout-Petit, Etudes de cas en statistique, 28h, L3 MASS (applied mathematics), Université Bordeaux Segalen, France.

Licence : A. Gégout-Petit, Econométrie et séries chronologiques, 24h, L3 MASS (applied mathematics), Université Bordeaux Segalen, France.

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

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

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

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

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

Master : F. Dufour, 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.

Master : B. de Saporta, Processus aléatoires en finance 30h ETD, M1, université de Bordeaux, France

Master : B. de Saporta, Finance en temps continu, 10h ETD, M2, université de Bordeaux, France

Master : B. de Saporta, Finance en temps discret, 29h ETD, M2, université de Bordeaux, France

Master : B. de Saporta, Processus de Markov, 25h ETD, M2, université de Bordeaux, France

Master : A. Gégout-Petit, Analyse de variance, 36h, M1, université Bordeaux, France.

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

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

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

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

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

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

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

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

Master: J. Saracco, Sampling techniques and experimental designs, 25h, M2, Master “Ingénierie Mathématique, Statistique et Economique”, the University of Bordeaux, France

8.6.2. Supervision

HdR : Anne Gégout-Petit, "Contribution à la statistique des processus : modélisation et applications, Université Bordeaux 2, 19 novembre 2012

PhD : Adrien Brandejski, Méthodes numériques pour les Processus Markoviens Déterministes par Morceaux,Université Bordeaux 1, 2 juillet 2012, supervised by F. Dufour and B. de Saporta

PhD in progress : Azaïs Romain, Inférence des processus Markoviens déterministe par morceaux , juillet 2013, supervised by François Dufour and Anne Gégout-Petit

PhD in progress : Camille Baysse, Analyse et optimisation de la fiabilité d’un équipement opto-électronique équipé de HUMS, novembre 2013, supervised by Anne Gégout-Petit and Jérôme Saracco

PhD in progress : Laurent Vezard, "Classification de signaux EEG et synthèse de paramètres musicaux par algorithme évolutionnaire", University of Bordeaux 1, supervised by M. Chavent and P. Legrand.

PhD in progress : Raphaël Coudret, Modélisation statistique de données acquises à haute fréquence : application en environnement et génétique, University of Bordeaux 1, supervised by J. Saracco and G.Durrieu.

PhD in progress : Karim Claudio, Un outil d’aide à la maîtrise des pertes dans les réseaux d’eau potable : mise en place d’un modèle de fuite multi-état en secteur hydraulique instrumenté , University of Bordeaux 1, supervised by J. Saracco and V. Couallier.

PhD in progress : Amaury Labenne, Approche Statistique du diagnostic territorial par la notion de qualité de vie, University of Bordeaux 1, supervised by M. Chavent, J. Saracco and V. Kuentz.

PhD in progress : Isabelle Charlier, Optimal quantization applied to conditional quantile estimation, University of Bordeaux 1 and Université Libre de Bruxelle, supervised by J. Saracco and D. Paindaveine.

8.6.3. Juries

B. de Saporta was a member of the selection committee for an assistant professor position (MdC 26) at University Montesquieu Bordeaux IV.

F. Dufour was referee of the PhD dissertation of C. Illand at Université Paris 6.

F. Dufour was referee of the PhD dissertation of Ariane Lorton at Université Technologique de Troyes.

J. Saracco was referee of the PhD dissertation of Adriana Cucu Gogonel at Université Paris 5.

J. Saracco was president of jury of the PhD of Adriana Christophe Denis at Université Paris 5.

J. Saracco was referee of the PhD dissertation of Adriana Cucu Gogonel at Université Paris 5.

J. Saracco was member of the juries for the HDR of Charles Bouveyron (Université Paris 1) and Julien Jacques (Université de Lille 1).

9. 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 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^{o} 5, p. 1607-1637.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[9] A. BRANDEJSKY. *Méthodes numériques pour les processus markoviens déterministes par morceaux*, Université Bordeaux 1, July 2012, http://tel.archives-ouvertes.fr/tel-00733731.

[10] A. GEGOUT-PETIT. *Contribution à la statistique des processus : modélisation et applications*, Université Sciences et Technologies - Bordeaux I, November 2012, Habilitation à Diriger des Recherches, http://hal.inria.fr/tel-00762189.

Articles in International Peer-Reviewed Journals

[11] 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, vol. 41, p. 17-36 [*DOI :* 10.1016/J.JKSS.2011.05.005], http://hal.inria.fr/hal-00642653.

[12] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Numerical method for expectations of piecewise-determistic Markov processes*, in "Communications in Applied Mathematics & Computational Science", 2012, vol. 7, n^{o} 1, p. 63-104, 39 pages, http://hal.inria.fr/hal-00617816.

[13] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Numerical methods for the exit time of a piecewise-deterministic Markov process*, in "Advances in Applied Probability", 2012, vol. 44, n^{o} 1, p. 196-225, http://hal.inria.fr/hal-00546339.

[14] M. CHAVENT, V. KUENTZ, B. LIQUET, J. SARACCO. *ClustOfVar: An R Package for the Clustering of Variables*, in "Journal of Statistical Software", 2012, vol. 50, n^{o} 13, p. 1-16, http://hal.inria.fr/hal-00764429.

[15] M. CHAVENT, V. KUENTZ-SIMONET, J. SARACCO. *Orthogonal rotation in PCAMIX*, in "Advances in Data Analysis and Classification", 2012, vol. 6, p. 131-146 [*DOI :* 10.1007/s11634-012-0105-3], http://hal.inria.fr/hal-00764427.

[16] F. DUFOUR, O. COSTA. *Singularly Perturbed Discounted Markov Control Processes in a General State Space*, in "SIAM Journal of control and optimization", 2012, vol. 50, n^{o} 2, p. 720-747, http://hal.inria.fr/hal-00759715.

[17] F. DUFOUR, M. HORIGUCHI, A. PIUNOVSKIY. *The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach*, in "Advances in Applied Probability", 2012, vol. 44, n^{o} 3, p. 774-793, http://hal.inria.fr/hal-00759717.

[18] F. DUFOUR, T. PRIETO-RUMEAU. *Approximation of Markov Decision Processes with General State Space*, in "Journal of Mathematical Analysis and applications", 2012, vol. 388, n^{o} 2, p. 1254-1267, http://hal.inria.fr/hal-00648223.

[19] 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", 2012, vol. 84, n^{o} 1, p. 57-78 [*DOI :* 10.1080/17442508.2011.580347], http://hal.inria.fr/hal-00648221.

[20] A. GÉGOUT-PETIT, R. AZAIS, 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.

[21] J. JOSSE, M. CHAVENT, B. LIQUET, F. HUSSON. *Handling Missing Values with Regularized Iterative Multiple Correspondence Analysis*, in "Journal of Classification", 2012, vol. 29, n° 1, p. 91-116, http://hal.inria.fr/hal-00763227.

[22] B. DE SAPORTA, F. DUFOUR. *Numerical method for impulse control of Piecewise Deterministic Markov Processes*, in "Automatica", 2012, vol. 48, p. 779-793, http://hal.inria.fr/hal-00541413.

[23] B. DE SAPORTA, F. DUFOUR, H. ZHANG, C. ELEGBEDE. *Optimal stopping for predictive maintenance of a structure subject to corrosion*, in "Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability", 2012, vol. 226, n° 2, p. 169-181, http://hal.inria.fr/hal-00554759.

[24] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Asymmetry tests for Bifurcating Auto-Regressive Processes with missing data*, in "Statistics and Probability Letters", 2012, vol. 82, n° 7, p. 1439-1444, http://hal.inria.fr/hal-00662129.

Articles in National Peer-Reviewed Journals

[25] 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 "Revue des Nouvelles Technologies de l'Information", 2012, http://hal.inria.fr/hal-00643438.

Invited Conferences

[26] B. DE SAPORTA, A. BRANDEJSKY, F. DUFOUR. *Optimal stopping for partially observed piecewise deterministic markov processes*, in "XIème Colloque Franco-Roumain de Mathématiques Appliquées", Bucarest, Roumanie, 2012, http://hal.inria.fr/hal-00755119.

[27] F. DUFOUR, M. HORIGUCHI, A. PIUNOVSKIY. *The expected total cost criterion for Markov decision processes under constraints : a convex analytic approach*, in "EURO 2012", Lituanie, 2012, http://hal.inria.fr/hal-00759727.

[28] F. DUFOUR, T. PRIETO-RUMEAU. *Linear Programming Approximations of Constrained Markov Decision Processes*, in "XIème Colloque Franco-Roumain de Mathématiques Appliquées,", Roumanie, 2012, http://hal.inria.fr/hal-00759729.

[29] J. SARACCO. *A sliced inverse regression approach for block-wise evolving data streams*, in "StatLearn 2012", Lille, France, April 2012, http://hal.inria.fr/hal-00768451.

International Conferences with Proceedings

[30] C. BAYSSE, A. GÉGOUT-PETIT, D. BIHANNIC, M. PRENAT, J. SARACCO. *Detection of a degraded operating mode of optronic equipment using Hidden Markov Model.*, in "PSAM 11 / ESREL 2012", Finlande, June 2012, 6, http://hal.inria.fr/hal-00762227.

[31] P. BRITO, M. CHAVENT. *Divisive Monothetic Clustering for Interval and Histogram-valued Data*, in "ICPRAM 2012 - 1st International Conference on Pattern Recognition Applications and Methods", Portugal, 2012, p. 229-234, http://hal.inria.fr/hal-00763215.

[32] B. DE SAPORTA, F. DUFOUR, H. ZHANG. *Predictive maintenance for the heated hold-up tank*, in "PSAM11-ESREL12", Helsinki, Finlande, 2012, http://hal.inria.fr/hal-00755078.

[33] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Analyse asymptotique des processus autoregressifs de bifurcation avec données manquantes*, in "44èmes Journées de Statistique", Bruxelles, Belgique, 2012, http://hal.inria.fr/hal-00755108.

[34] A. GÉGOUT-PETIT, M. TOUZET-CORTINA, M. PUIGGALI, A. BEN ABDESSALEM, R. AZAÏS. *Modelling of fatigue crack propagation using Piecewise Deterministic Markov Processes*, in "3rd International Conference on Engineering Optimization", Brésil, June 2012, 6, http://hal.inria.fr/hal-00762214.

[35] A. NÈGRE, O. MARCEAU, D. LANEUVILLE, H. ZHANG, B. DE SAPORTA, F. DUFOUR. *Stochastic control for underwater optimal trajectories*, in "IEEE Aerospace conference", Big Sky, États-Unis, 2012 [*DOI :* 10.1109/AERO.2012.6187191], http://hal.inria.fr/hal-00755115.

[36] H. ZHANG, B. DE SAPORTA, F. DUFOUR, G. DELEUZE. *Dynamic reliability : towards efficient simulation of the availability of a feedwater control system*, in "NPIC-HMIT", San Diego, États-Unis, 2012, p. 714-723, http://hal.inria.fr/hal-00755070.

National Conferences with Proceeding

[37] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Estimation non-paramétrique du taux de saut d'un processus de renouvellement marqué non-homogène*, in "44èmes Journées de Statistique", Bruxelles, Belgique, May 2012, http://hal.inria.fr/hal-00762912.

[38] C. BAYSSE, D. BIHANNIC, B. DE SAPORTA, A. GÉGOUT-PETIT, M. PRENAT, J. SARACCO. *Modèle de markov caché pour la détection d'un mode de fonctionnement dégradé d'un équipement optronique*, in "Lambda-Mu 18", Tours, France, 2012, communication 3D-1, http://hal.inria.fr/hal-00755131.

[39] H. ZHANG, B. DE SAPORTA, F. DUFOUR, G. DELEUZE. *Fiabilité dynamique : simulation d'un système de régulation du niveau d'eau dans un générateur de vapeur*, in "Lambda-Mu 18,", La Rochelle, France, 2012, Communication 7D-1, http://hal.inria.fr/hal-00755125.

Conferences without Proceedings

[40] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *A nonparametric estimator of the jump rate for a general class of marked renewal processes*, in "Sexto encuentro de Estadística Matemática BoSanTouVal", Castro Urdiales, Espagne, June 2012, http://hal.inria.fr/hal-00762913.

[41] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Estimation non-paramétrique de la loi conditionnelle des temps inter-sauts d'un PDMP*, in "Journées MAS 2012 Clermont-Ferrand", Clermont-Ferrand, France, August 2012, http://hal.inria.fr/hal-00762914.

[42] M. CHAVENT, V. KUENTZ-SIMONET, Z. LAKATOS, J. SARACCO. *Rotation orthogonale en ACP de données mixtes. Le package PCAmixdata et une application en sociologie culturelle.*, in "1ères Rencontres R", Bordeaux, France, July 2012, http://hal.inria.fr/hal-00717515.

[43] R. COUDRET, G. DURRIEU, J. SARACCO. *Une interface graphique pour analyser des données distantes sous R*, in "1ères Rencontres R", Bordeaux, France, July 2012, Ce travail présente une méthode pour implémenter des algorithmes d'analyse de données lorsque celles-ci sont stockées sur une machine distante et pour créer une interface graphique facilitant l'utilisation de ces algorithmes. Une application à des estimations de densité dans un contexte biologique est fournie., http://hal.inria.fr/hal-00717552.

[44] A. GÉGOUT-PETIT, L. GUERIN-DUBRANA, J.-C. LABROUSSE, A. LABENNE, P. REY, S. BASTIEN. *Statistical Analysis of Grapevine Mortality Associated with Esca or Eutypa Dieback Foliar Expression*, in "8th International Workshop on Grapevine Trunk Diseases", Valencia, Espagne, June 2012, http://hal.inria.fr/hal-00762208.

[45] V. KUENTZ-SIMONET, S. LYSER, M. CHAVENT, J. SARACCO, J. CANDAU, P. DEUFFIC. *La prise en compte de l'environnement par les agriculteurs : une analyse avec le package "ClustOfVar"*, in "1ères Rencontres R", Bordeaux, France, July 2012, http://hal.inria.fr/hal-00717496.

[46] A. LABENNE, F. BONNEU, M. CHAVENT, A. GEGOUT-PETIT, L. GUERIN-DUBRANA. *Test de la vraisemblance entre deux motifs de points*, in "1ères Rencontres R", Bordeaux, France, July 2012, http://hal.inria.fr/hal-00717541.

[47] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA, J. CLAUZEL. *Classification of EEG signals by an evolutionary algorithm*, in "COMPSTAT 2012", Limassol, Chypre, August 2012, http://hal.inria.fr/hal-00757270.

[48] L. VEZARD, P. LEGRAND, M. CHAVENT, F. FAITA-AINSEBA. *Sélection de variables par algorithme génétique pour l'étude d'états de vigilance*, in "Journée Évolutionnaire Thématique, 23éme édition", Paris, France, November 2012, http://hal.inria.fr/hal-00759420.

Scientific Books (or Scientific Book chapters)

[49] J.-F. AUBRY, G. BABYKINA, N. BRINZEI, S. MEDJAHER, A. BARROS, C. BÉRENGUER, A. GRALL, Y. LANGERON, D. NGOC NGUYEN, G. DELEUZE, B. DE SAPORTA, F. DUFOUR, H. ZHANG. *Projet APPRODYN : approches de la fiabilité dynamique pour modéliser des systèmes critiques*, in "Supervision, surveillance et sûreté de fonctionnement des grands systèmes", N. MATTA, Y. VANDENBOOMGAERDE, J. ARLAT (editors), Traité Information, Commande, Communication, IC2, Hermès Science Publications, March 2012, p. 181-222, http://hal.inria.fr/hal-00675333.

[50] J.-F. AUBRY, G. BABYKINA, N. BRINZEI, S. MEDJAHER, A. BARROS, C. BÉRENGUER, A. GRALL, Y. LANGERON, D. NGOC NGUYEN, G. DELEUZE, B. DE SAPORTA, F. DUFOUR, H. ZHANG. *The APPRODYN project: dynamic reliability approaches to modeling critical systems*, in "Supervision and Safety of Complex Systems", N. MATTA, Y. VANDENBOOMGAERDE, J. ARLAT (editors), Wiley-ISTE, August 2012, p. 181-222, Chapter 8, http://hal.inria.fr/hal-00762917.

[51] O. COSTA, F. DUFOUR. *Continuous Control of Piecewise Deterministic Markov Processes with Long Run Average Cost*, in "Stochastic Processes, Finance and Control. A Festschrift in Honor of Robert J. Elliott.", World Scientific, 2012, p. 120-154, http://hal.inria.fr/hal-00759726.

[52] F. DUFOUR, T. PRIETO-RUMEAU. *Approximation of Infinite Horizon Discounted Cost Markov Decision Processes*, in "Optimization, Control, and Applications of Stochastic Systems", Birkhäuser, 2012, p. 59-76, http://hal.inria.fr/hal-00759719.

Research Reports

[53] J.-F. AUBRY, G. BABYKINA, A. BARROS, N. BRINZEI, G. DELEUZE, B. DE SAPORTA, F. DUFOUR, Y. LANGERON, H. ZHANG. *Rapport final du projet APPRODYN : APPROches de la fiabilité DYNamique pour modéliser des systèmes critiques*, Inria, March 2012, n^{o} SURFDIAG, ISET, http://hal.inria.fr/hal-00740181.

[54] H. ZHANG, B. DE SAPORTA, F. DUFOUR. *Contrôle optimal stochastique, application à l'optimisation de trajectoire en 3d*, Inria, 2012, Rapport technique de CQFD pour DCNS, http://hal.inria.fr/hal-00755134.

Other Publications

[55] R. AZAÏS. *A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process*, 2012, http://hal.inria.fr/hal-00759065.

[56] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the conditional distribution of the inter-jumping times for piecewise-deterministic Markov processes*, 2012, http://hal.inria.fr/hal-00759064.

[57] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes*, 2012, http://hal.inria.fr/hal-00759062.

[58] C. BAYSSE, D. BIHANNIC, A. GÉGOUT-PETIT, M. PRENAT, J. SARACCO. *Hidden Markov Model for the detection of a degraded operating mode of optronic equipment*, 2012, 11 pages, 9 figures, http://hal.inria.fr/hal-00763963.

[59] B. BERCU, T. MONG NGOC NGUYEN, J. SARACCO. *On the asymptotic behavior of the Nadaraya-Watson estimator associated with the recursive SIR method*, 2012, http://hal.inria.fr/hal-00673832.

[60] A. BRANDEJSKY, B. DE SAPORTA, F. DUFOUR. *Optimal stopping for partially observed piecewise-deterministic Markov processes*, 2012, http://hal.inria.fr/hal-00755052.

[61] M. CHAVENT, S. GIRARD, V. KUENTZ, B. LIQUET, T. M. N. NGUYEN, J. SARACCO. *A sliced inverse regression approach for data stream*, http://hal.inria.fr/hal-00688609.

[62] R. COUDRET, G. DURRIEU, J. SARACCO. *Comparison of kernel density estimators with assumption on number of modes*, 2012, http://hal.inria.fr/hal-00748790.

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

[64] B. DE SAPORTA, A. GÉGOUT-PETIT, L. MARSALLE. *Random coefficients bifurcating autoregressive processes*, 2012, http://hal.inria.fr/hal-00702357.

[65] B. DE SAPORTA, A. G. PETIT, L. MARSALLE. *Statistical study of asymmetry in cell lineage data*, 2012, http://hal.inria.fr/hal-00702359.

[66] B. DE SAPORTA, H. ZHANG. *Predictive maintenance for the heated hold-up tank*, 2012, arXiv admin note: substantial text overlap with arXiv:1101.1740, http://hal.inria.fr/hal-00755056.

References in notes

[67] 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.

[68] T. AVEN, U. JENSEN. *Stochastic models in reliability*, Applications of Mathematics, Springer, New York, 1999, vol. 41.

[69] B. BERCU, B. DE SAPORTA, A. GÉGOUT-PETIT. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach*, in "Electronic Journal of Probability", 2009, vol. 14, p. 2492–2526.

[70] V. BLANDIN. *Asymptotic results for bifurcating random coefficient autoregressive processes*, 2012, ArXiv 1204.2926.

[71] Q. M. BUI, R. M. HUGGINS. *Inference for the random coefficients bifurcating autoregressive model for cell lineage studies*, in "J. Statist. Plann. Inference", 1999, vol. 81, n^{o} 2, p. 253–262.

[72] R. COWAN, R. G. STAUDTE. *The bifurcating autoregressive model in cell lineage studies*, in "Biometrics", 1986, vol. 42, p. 769–783.

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

[74] J.-F. DELMAS, L. MARSALLE. *Detection of cellular aging in a Galton-Watson process*, in "Stoch. Process. and Appl.", 2010, vol. 120, p. 2495–2519.

[75] J. DEVOOGHT. *Dynamic reliability*, Advances in nuclear science and technology, Chapman and Hall, Berlin, 1997.

[76] 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.

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

[78] M. DUFLO. *Random iterative models*, Applications of Mathematics, Springer-Verlag, Berlin, 1997, vol. 34, xviii+385.

[79] F. DUFOUR, Y. DUTUIT. *Dynamic reliability: A new model*, in "Proceedings of ESREL 2002 Lambda-Mu 13 Conference", 2002, p. 350–353.

[80] A. GRALL, C. BÉRENGUER, L. DIEULLE. *A conditional-based maintenance policy for a stochastically deteriorating system*, in "Reliability Engineering and System Safety", 2002, vol. 76, n^{o} 2, p. 167-180.

[81] U. S. GUGERLI. *Optimal stopping of a piecewise-deterministic Markov process*, in "Stochastics", 1986, vol. 19, n^{o} 4, p. 221–236.

[82] J. GUYON. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*, in "Ann. Appl. Probab.", 2007, vol. 17, n^{o} 5-6, p. 1538–1569.

[83] R. M. HUGGINS, I. V. BASAWA. *Inference for the extended bifurcating autoregressive model for cell lineage studies*, in "Aust. N. Z. J. Stat.", 2000, vol. 42, n^{o} 4, p. 423–432.

[84] R. M. HUGGINS, I. V. BASAWA. *Extensions of the bifurcating autoregressive model for cell lineage studies*, in "J. Appl. Probab.", 1999, vol. 36, n^{o} 4, p. 1225–1233.

[85] R. M. HUGGINS. *Robust inference for variance components models for single trees of cell lineage data*, in "Ann. Statist.", 1996, vol. 24, n^{o} 3, p. 1145–1160.

[86] R. M. HUGGINS, R. G. STAUDTE. *Variance Components Models for Dependent Cell Populations*, in "J. AMS", 1994, vol. 89, n^{o} 425, p. 19–29.

[87] 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.

[88] 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.

[89] M. MARSEGUERRA, E. ZIO. *The cell-to-boundary method in Monte Carlo-based dynamic PSA*, in "Reliability Engineering and System Safety", 1995, vol. 48, n^{o} 3, p. 199-204.

[90] M. MARSEGUERRA, E. ZIO. *Monte Carlo approach to PSA for dynamic process systems*, in "Reliability Engineering and System Safety", 1996, vol. 52, n^{o} 3 SPEC. ISS., p. 227-241.

[91] D. F. NICHOLLS, B. G. QUINN. *Random coefficient autoregressive models: an introduction*, Lecture Notes in Statistics, Springer-Verlag, New York, 1982, vol. 11, Lecture Notes in Physics, 151.

[92] H. PHAM, W. RUNGGALDIER, A. SELLAMI. *Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation*, in "Monte Carlo Methods Appl.", 2005, vol. 11, n^{o} 1, p. 57–81.

[93] E. STEWART, R. MADDEN, F. TADDEI. *Aging and death in an organism that reproduces by morphologically symmetric division*, in "PLoS Biol.", 2005, vol. 3, n^{o} 2, e45.

[94] H. ZHANG, F. DUFOUR, Y. DUTUIT, K. GONZALEZ. *Piecewise deterministic Markov processes and dynamic reliability*, in "Proc. of the Inst. of Mechanical Engineers, Part O: Journal of Risk and Reliability", 2008, vol. 222, n^{o} 4, p. 545–551.

[95] J. ZHOU, I. V. BASAWA. *Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors*, in "J. Time Ser. Anal.", 2005, vol. 26, n^{o} 6, p. 825–842.

[96] J. VAN NOORTWIJK. *A survey of the application of gamma processes in maintenance*, in "Reliability Engineering and System Safety", 2009, vol. 94, n^{o} 1, p. 2-21.