RESEARCH CENTRE

Inria Center at the University of Bordeaux

IN PARTNERSHIP WITH:

Université de Bordeaux, Institut Polytechnique de Bordeaux, Naval Group, CNRS

2022 ACTIVITY REPORT

Project-Team ASTRAL

Advanced StatisTical infeRence And controL

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

DOMAIN

Applied Mathematics, Computation and Simulation

THEME

Stochastic approaches

Inria

Contents

Project-Team ASTRAL

Creation of the Project-Team: 2021 January 01

Keywords

Computer sciences and digital sciences

A3.4. – Machine learning and statistics

A3.4.1. – Supervised learning

A3.4.2. – Unsupervised learning

A3.4.3. – Reinforcement learning

A3.4.4. – Optimization and learning

A3.4.5. – Bayesian methods

A3.4.6. – Neural networks

A3.4.7. – Kernel methods

A3.4.8. – Deep learning

A6.1.2. – Stochastic Modeling

A6.1.3. – Discrete Modeling (multi-agent, people centered)

A6.2.2. – Numerical probability

A6.2.3. – Probabilistic methods

A6.2.4. – Statistical methods

A6.2.6. – Optimization

A6.3.3. – Data processing

A6.3.4. – Model reduction

A6.3.5. – Uncertainty Quantification

A6.4. – Automatic control

A6.4.1. – Deterministic control

A6.4.2. – Stochastic control

A6.4.3. – Observability and Controlability

A6.4.4. – Stability and Stabilization

A6.4.5. – Control of distributed parameter systems

A6.4.6. – Optimal control

A8.2.2. – Evolutionary algorithms

A8.11. – Game Theory

A9.2. – Machine learning

A9.3. – Signal analysis

A9.6. – Decision support

A9.7. – AI algorithmics

Other research topics and application domains

B1.1.2. – Molecular and cellular biology

B1.2.3. – Computational neurosciences

B2.5.1. – Sensorimotor disabilities

B4.2.1. – Fission

1 Team members, visitors, external collaborators

Research Scientists

- Pierre Del Moral [INRIA, Senior Researcher, HDR]
- Emma Horton [INRIA, Researcher]
- Dann Laneuville [Engineer NAVAL GROUP, Industrial member]
- Olivier Marceau [Engineer NAVAL GROUP, Industrial member]
- Adrien Negre [Engineer NAVAL GROUP, Industrial member]

Faculty Members

- François Dufour [Team leader, BORDEAUX INP, Professor, HDR]
- Marie Chavent [UNIV BORDEAUX, Professor, HDR]
- Alexandre Genadot [UNIV BORDEAUX, Associate Professor]
- Pierrick Legrand [UNIV BORDEAUX, Associate Professor, HDR]
- Jérôme Saracco [BORDEAUX INP, Professor, HDR]

Post-Doctoral Fellow

- Hadrien Lorenzo [INRIA]

PhD Students

- Luc De Montella [NAVAL GROUP, CIFRE, from Oct 2022]
- Mariette Dupuy [Inria/IHU Lyric]
- Romain Namyst [INRIA]
- Valentin Portmann [Inria, from Oct 2022]
- Tara Vanhatalo [OROSYS]

Technical Staff

- Enzo Iglesis [Inria, Engineer, from Nov 2022]

Interns and Apprentices

- Luc De Montella [NAVAL GROUP, from Apr 2022 until Sep 2022]

Administrative Assistants

- Audrey Plaza [INRIA]
- Rima Soueidan [Inria, from Sep 2022]

Visiting Scientists

- Bertrand Cloez [INRAE, from Jun 2022]
- Alexander Cox [UNIV BATH, from Oct 2022]
- Alastair James Crossley [UNIV BATH, from Sep 2022]
- Oswaldo Do Valle Costa [UNIV SAO PAULO, from Sep 2022]
- Sophie Hautphenne [UNIV MELBOURNE, from Jul 2022 until Jul 2022]
- Ajay Jasra [KAUST, from Nov 2022]
- Oscar Key [UCL, from Nov 2022]
- Tony Lelièvre [ENPC, from Jul 2022, HDR]
- Bastien Mallein [UNIV PARIS XIII, from Sep 2022]
- Pierre Kristoffer Nyquist [INSTITUT KTH, from Oct 2022]
- Ellen Powell [UNIV DURHAM, from Jul 2022]
- Tomas Prieto Rumeau [UNED, from Oct 2022]
- Alexander Watson [UCL, from Nov 2022]

2 Overall objectives

2.1 Outline of the research project

The highly interconnected contemporary world is faced with an immense range of serious challenges in statistical learning, engineering and information sciences which make the development of statistical and stochastic methods for complex estimation problems and decision making critical. The most significant challenges arise in risk analysis, in environmental and statistical analysis of massive data sets, as well as in defense systems. From both the numerical and the theoretical viewpoints, there is a need for unconventional statistical and stochastic methods that go beyond the current frontier of knowledge.

Our approach to this interdisciplinary challenge is based on recent developments in statistics and stochastic computational methods. We propose a work programme which will lead to significant breakthroughs in fundamental and applied mathematical research, as well as in advanced engineering and information sciences with industrial applications with a particular focus on defence applications, in collaboration with Naval Group.

Many real-world systems and processes are dynamic and essentially random. Examples can be found in many areas like communication and information systems, biology, geophysics, finance, economics, production systems, maintenance, logistics and transportation. These systems require dynamic and stochastic mathematical representations with discrete and/or continuous state variables in possibly infinite dimensional space. Their dynamics can be modeled in discrete or continuous time according to different time scales and are governed by different types of processes such as stochastic differential equations, piecewise deterministic processes, jump-diffusion processes, branching and mean field type interacting processes, reinforced processes and self-interacting Markov processes, to name a few. Our interdisciplinary project draws knowledge from information science, signal processing, control theory, statistics and applied probability including numerical and mathematical analysis. The idea is to work across these scientific fields in order to enhance their understandings and to offer an original theory or concept.

Our group mainly focuses on the development of advanced statistical and probabilistic methods for the analysis and the control of complex stochastic systems, as outlined in the following three topics.

- ***Statistical and Stochastic modeling:*** Design and analysis of realistic and tractable statistical and stochastic models, including measurement models, for complex real-life systems taking into account various random phenomena. Refined qualitative and quantitative mathematical analysis of the stability and the robustness of statistical models and stochastic processes.

- ***Estimation/Calibration:*** Theoretical methods and advanced computational methodologies to estimate the parameters and the random states of the model given partial and noisy measurements as well as statistical data sets. Refined mathematical analysis of the performance and the convergence of statistical and stochastic learning algorithms.

- ***Decision and Control:*** Theoretical methods and advanced computational methodologies for solving regulation and stochastic optimal control problems, including optimal stopping problems and partially observed models. Refined mathematical analysis of the long time behavior and the robustness of decision and control systems.

These three items are by no means independent.

- Regarding the interdependence between the modeling aspects and the estimation/calibration/control aspects, it must be emphasized that when optimizing the performance of a partially observed/known stochastic system, the involved mathematical techniques will heavily depend on the underlying mathematical characteristics and complexity of the model of the state process and the model of the observation process. The main difficulty here is to find a balance between complexity and feasibility/solvability. The more sophisticated a model is, the more complicated the statistical inference and optimization problems will be to solve.

- The interdependence that arises between estimation/calibration and the optimal control can be summarised as follows. When the decision-maker has only partial information on the state process, it is necessary to assume that the admissible control policies will depend on the filtration generated by the observation process. This is a particularly difficult optimisation problem to solve. Roughly speaking, by introducing the conditional distribution of the state process, the problem can then be reformulated in terms of a fully observed control problem. This leads to a separation of estimation and control principle, i.e. the estimation step is carried out first and then the optimisation. The price to be paid for this new formulation is an enlarged state space of infinite dimension. More precisely, in addition to the observable part of the state, a probability distribution enters the new state space which defines the conditional distribution of the unobserved part of the state given the history of the observations.

Solving such global optimization problems remain an open problem and is recognized in the literature as a very difficult challenge to meet.

One of the fundamental challenges we will address is to develop estimation/calibration and optimal control techniques related to general classes of stochastic processes in order to deal with real-world problems. Our research results will combine, mathematical rigour (through the application of advanced tools from probability, statistics, measure theory, functional analysis and optimization) with computational efficiency (providing accurate and applicable numerical methods with a refined analysis of the convergence). Thus, the results that we will obtain in this research programme will be of interest to researchers in the fields of stochastic modeling, statistics and control theory both for the theoretical and the applied communities. Moreover, the topics studied by Naval Group, such as target detection, nonlinear filtering, multi-object tracking, trajectory optimization and navigation systems, provide a diverse range of application domains in which to implement and test the methodologies we wish to develop.

The final goal is to develop a series of reliable and robust softwares dedicated to statistical and stochastic learning, as well as automated decision and optimal control processes. The numerical codes are required to be both accurate and fast since they are often elements of real time estimation and control loops in automation systems. In this regard, the research topics proposed by Naval Group will provide a natural framework for testing the efficiency and robustness of the algorithms developed by the team.

From our point of view, this collaboration between the INRIA project team and Naval Group offers new opportunities and strategies to design advanced cutting-edge estimation and control methodologies.

2.2 Approach and methodologies

The types of learning and control methodologies developed by the team differ in their approach as well as in the problems that they are intended to solve. They can be summarised by the following three sets of interdependent methodologies.

- ***Statistical learning:*** Regression, clustering, volume and dimensionality reduction, classification, data mining, training sets analysis, supervised and unsupervised learning, active and online learning, reinforcement learning, identification, calibration, Bayesian inference, likelihood optimisation, information processing and computational data modeling.

- ***Stochastic learning:*** Advanced Monte Carlo methods, reinforcement learning, local random searches, stochastic optimisation algorithms, stochastic gradients, genetic programming and evolutionary algorithms, interacting particle and ensemble methodologies, uncertainty propagation, black box inversion tools, uncertainty propagation in numerical codes, rare event and default tree simulation, nonlinear and high dimensional filtering, prediction and smoothing.

- ***Decision and control:*** Markov decision processes, piecewise deterministic Markov processes, stability, robustness, regulation, optimal stopping, impulse control, stochastic optimal control including partially observed problems, games, linear programming approaches, dynamic programming techniques.

All team members of the project work at the interface of the these three areas. This joint research project between INRIA and Naval Group is a natural and unprecedented opportunity to embrace and push the frontiers of the applied and theoretical sides of these research topics in a common research team.

Despite some recent advances, the design and the mathematical analysis of statistical and stochastic learning tools, as well as automated decision processes, is still a significant challenge. For example, since the mid-1970s nonlinear filtering problems and stochastic optimal control problems with partial observations have been the subject of several mathematical studies, however very few numerical solutions have been proposed in the literature.

Conversely, since the mid-1990s, there has been a virtual explosion in the use of stochastic particle methods as powerful tools in real-word applications of Monte Carlo simulation; to name a few, particle filters, evolutionary and genetic algorithms and ensemble Kalman filters. Most of the applied research in statistics, information theory and engineering sciences seems to be developed in a completely blind way with no apparent connections to the mathematical counterparts.

This lack of communication between the fields often produces a series of heuristic techniques often tested on reduced or toy models. In addition, most of these methodologies do not have a single concrete industrial application nor do they have any connection with physical problems.

As such, there exists a plethora of open mathematical research problems related to the analysis of statistical learning and decision processes. For instance, a variety of theoretical studies on particle algorithms, including particle filters and sequential Monte Carlo models are often based on ad-hoc and practically unrealistic assumptions for the kinds of complex models that are increasingly emerging in applications.

The aim of this project is to fill these gaps with an ambitious programme at the intersection of probability, statistics, engineering and information sciences.

One key advantage of the project is its interdisciplinary nature. Combining techniques from pure and applied mathematics, applied probability and statistics, as well as computer science, machine learning, artificial intelligence and advanced engineering sciences enables us to consider these topics holistically, in order to deal with real industrial problems in the context of risk management, data assimilation, tracking applications and automated control systems. The overarching aim of this ambitious programme is to make a breakthrough in both the mathematical analysis and the numerical aspects of statistical learning and stochastic estimation and control.

2.3 Innovation and industrial transfer

Fundamentally, our team is not driven by a single application. The reasons are three-fold. Firstly, the robustness and transferability of our approaches means that the same statistical or stochastic learning algorithms can be used in a variety of application areas. On the other hand every application domain offers a series of different perspectives that can be used to improve the design and performances of our techniques and algorithms. Last but not least, industrial applications, including those that arise in defence, require specific attention. As such, we use a broad set of stochastic and statistical algorithms to meet these demands.

This research programme is oriented towards concrete applications with significant potential industrial transfers on three central problems arising in engineering and information and data sciences, namely, risk management and uncertainty propagation, process automation, and data assimilation, tracking and guidance. Our ultimate goal is to bring cutting edge algorithms and advanced statistical tools to industry and defence. The main application domains developed by the team are outlined below:

- ***Risk management and uncertainty propagation:*** Industrial and environmental risks, fault diagnostics, phase changes, epidemiology, nuclear plants, financial ruin, systemic risk, satellite debris collisions.

- ***Process automation:*** Production maintenance and manufacturing planing, default detection, integrated dynamics and control of distributed dynamical systems, multi-object coordination, automatic tuning of cochlear implants, classification of EGG signals.

- ***Data assimilation, tracking and guidance:*** Target detection and classification, nonlinear filtering and multi-object tracking, multiple sensor fusions, motion planning, trajectory optimization, design of navigation systems.

The main objectives and challenges related to the three application domains discussed above will be developed in section 4. The latter application domain will be developed in collaboration with Naval Group. The reader is refereed to section 4.1 for a description of this collaboration and to sections 3.2 and 3.3 for the theoretical aspects that will be carried out by the team in relation to these topics. Specific details on the particular techniques used to tackle the estimation and tracking problems in the context of the collaboration with Naval Group will remain confidential.

3 Research program

This section describes the different challenges we intend to address in the theoretical and numerical aspects of statistical/stochastic learning and optimal control. It will be difficult to convey the full complexity of the various topics and to provide a complete overview through a detailed timetable. Nevertheless, we will explain our motivation and why we think it is imperative to address these challenges. We will also highlight the technical issues inherent to these challenges, as well as the difficulties we might expect.

We are confident that the outcomes of this scientific project will lead to significant breakthroughs in statistical/stochastic learning and optimal control with a special emphasis on applications in the defence industry in collaboration with Naval Group. In this respect, we would like to quote Hervé Guillou, CEO of Naval Group, on the occasion of the signing of the partnership agreement between INRIA and Naval Group on December 10, 2019: *"This partnership will enable Naval Group to accelerate its innovation process in the fields of artificial intelligence, intelligence applied to cyber and signal processing. This is a necessity given the French Navy's need for technological superiority in combat and the heightened international competition in the naval defence field..."*

One of our greatest achievements would undoubtedly be to meet these challenges with Naval Group, particularly those related to the fields of statistical/stochastic learning and control. We could not dream of a better outcome for our project.

3.1 Statistical learning

<u>Permanent researchers:</u> M. Chavent, P. Del Moral, F. Dufour, A. Genadot, E. Horton, P. Legrand, J. Saracco.

Regarding statistical learning, some of the objectives of the team is to develop dimension reduction models, data visualization, non-parametric estimation methods, genetic programming and artificial evolution. These models/methodologies provide a way to understand and visualize the structure of complex data sets. Furthermore, they are important tools in several different areas of research, such as data analysis and machine learning, that arise in many applications in biology, genetics, environment and recommendation systems. Of particular interest is the analysis of classification and clustering approaches and semi-parametric modeling that combines the advantages of parametric and non-parametric models, amongst others. One major challenge is to tackle both the complexity and the quantity of data when working on real-world problems that emerge in industry or other scientific fields in academia. Of particular interest is to find ways to handle high-dimensional data with irrelevant and redundant information.

Another challenging task is to take into account successive arrivals of information (data stream) and to dynamically refine the implemented estimation algorithms, whilst finding a balance between the estimation precision and the computational cost. One potential method for this is to project the available information into suitably chosen lower dimensional spaces.

For regression models, sliced inverse regression (SIR) and related approaches have proven to be highly efficient methods for modeling the link between a dependent variable (which can be multidimensional) and multivariate covariates in several frameworks (data stream, big data, etc.). The underlying regression model is semi-parametric (based on a single index or on multiple indices that allow dimension reduction). Currently, these models only deal with quantitative covariates. One of the team's goals is to extend these regression models to mixed data, i.e. models dealing with quantitative and categorical covariates. This generalization would allow one to propose discriminant analysis to deal with mixed data. Extension of sparse principal component analysis (PCA) to mixed data is also another challenge. One idea is to take inspiration from the underlying theory and method of recursive SIR and SIR approaches for data stream in order to adapt them to commonly used statistical methods in multivariate analysis (PCA, discriminant analysis, clustering, etc.). The common aim of all these approaches is to estimate lower dimensional subspaces whilst minimizing the loss of statistical information. Another important aspect of data stream is the possible evolution in time of the underlying model: we would like to study break(s) detection in semi-parametric regression model, the breakdown being susceptible to appear in the parametric part or in the functional part of the regression model. The question of selecting covariates in regression modelling when we deal with big data is a fundamental and difficult problem. We will address this challenge using genetic programming and artificial evolution. Several directions are possible: for instance, improve, via genetic algorithms, the exploration of the covariate space in closest submodel selection (CSS) method or study optimization problems that simultaneously take into account variable selection, efficiency of estimation and interpretability of the model. Another important question concerns the detection of outliers that will disturb the estimation of the model, and this is not an obvious problem to deal with when working with large, high dimensional data.

In multivariate data analysis, an objective of the team is to work on a new formulation/algorithm for group-sparse block PCA since it is always important to take into account group information when available. The advantage of the group-sparse block PCA is that, via the selection of groups of variables (based on the synthetic variables), interpretability of the results becomes easier. The underlying idea is to address the simultaneous determination of group-sparse loadings by block optimization, and the correlated problem of defining explained variance for a set of non-orthogonal components. The team is also interested in clustering of supervised variables, the idea being to construct clusters made up of variables correlated with each other, which are either well-linked or not-linked to the variable to be explained (which can be quantitative or qualitative).

Another way to study the links between variables is to consider conditional quantiles instead of conditional expectation as is the case in classical regression models. Indeed, it is often of interest to model conditional quantiles, particularly in the case where the conditional mean fails to take into account the impact of the covariates on the dependent variable. Moreover, the quantile regression function provides a much more comprehensive picture of the conditional distribution of a dependent variable than the conditional mean function. The team is interested in the non parametric estimation of conditional quantile estimation. New estimators based on quantization techniques have been introduced and studied in the literature for univariate conditional quantiles and multivariate conditional quantiles. However, there are still many open problems, such as combining information from conditional quantiles of different orders in order to refine the estimation of a conditional quantile of a given order.

Another topic of interest is genetic programming (GP) and Artificial Evolution. GP is an evolutionary computation paradigm for automatic program induction. GP has produced impressive results but there are still some practical limitations, including its high computational cost, overfitting and excessive code growth. Recently, many researchers have proposed fitness-case sampling methods to overcome some of these problems, with mixed results in several limited tests. Novelty Search (NS) is a unique approach towards search and optimization, where an explicit objective function is replaced by a measure of solution novelty. While NS has been mostly used in evolutionary robotics, the team would like to explore its usefulness in classic machine learning problems.

Another important objective of the team is to implement new R (Matlab/Python) packages or to enrich those existing in the literature with the methods we are going to develop in order to make them accessible to the scientific community.

With respect to our statistical learning research program, the objectives of the team can be divided into mid- and long-term works. Mid-term objectives focus on sparsity in SIR (via soft thresholding for instance) and group-sparse block PCA, the underlying idea being to make the selection of variables or blocks of variables in the regression model or in the data. Taking into account multi-block data in regression models via data-driven sparse partial least squares is also at the heart of our concerns. Coupling genetic algorithms and artificial evolution with statistical modeling issues is also planned. The team has several long-term projects associated with the notion of data stream. Many theoretical and practical problems arise from the possible evolution of the information contained in the data: break detection in the underlying model, balance between precision and computational cost. Another scientific challenge is to extend certain approaches such as SIR to the case of mixed data by incorporating the information provided by the qualitative variables in the associated low dimensional subspaces. Moreover, the team has already worked on clustering of variables for mixed data and the clustering of supervised variables is now planned. Finally the idea of combining information from conditional quantiles of different orders in order to refine the estimation of a given order conditional quantile is still relevant today. It should be noted that other research themes may appear or become a priority depending on the academic or industrial collaborations that may emerge during the next evaluation period.

Project-team positioning: Some topics of the INRIA project teams (STATIFY, CELESTE, MODAL, SEQUEL, CLASSIC) are close to the ASTRAL objectives such as non parametric view of high dimensional data, statistical/machine learning, model selection, clustering, sequential learning algorithms, or multivariate data analysis for complex data. While certain ASTRAL objectives are similar to those of these teams, our approaches are significantly different. For example, in multivariate data analysis of complex data including clustering, our team mainly focuses on a geometric approach for mixed data. We also consider the case of successive arrivals of information in SIR both from the theoretical and numerical point of view. Currently there is no direct competition between our team and other INRIA project teams. However, interactions between ASTRAL and other INRIA teams exist. For instance, ASTRAL and STATIFY collaborations are fruitful with common publications, in particular with S. Girard (STATIFY project team). In the field of multivariate data analysis, the team have interesting discussions with Agrocampus Ouest (Rennes, France) and with H.A.L. Kiers (Groningen University) on a mixed data approach for dimension reduction. Conditional and regression quantiles are very active research fields in France (University of Toulouse, Toulouse School of Economics, University of Montpellier) and around the world (ULB, Belgium; University of Illinois Urbana-Champaign, USA; Open University, UK; Brunel University, UK). The ASTRAL team has for the last four-year period collaborated with D. Paindaveine (ULB, Belgium). In the dimension reduction framework, there is a large international community in Europe, America or Asia working on SIR and related methods. However, to our knowledge, the ASTRAL team was the first to introduce importance of variables and recursive methods in SIR, and the first to adapt the SIR approach to data stream.

3.2 Stochastic learning

Permanent researchers: M. Chavent, P. Del Moral, F. Dufour, A. Genadot, E. Horton, D. Laneuville, P. Legrand, A. Nègre, J. Saracco, H. Zhang.

Stochastic particle methodologies have become one of the most active intersections between pure and applied probability theory, Bayesian inference, statistical machine learning, information theory, theoretical chemistry, quantum physics, financial mathematics, signal processing, risk analysis, and several other domains in engineering and computer sciences.

Since the mid-1990s, rapid developments in computer science, probability and statistics have led to new generations of interacting particle learning/sampling type algorithms, such as:

Particle and bootstrap filters, sequential Monte Carlo methods, self-interacting and reinforced learning schemss, sequentially interacting Markov chain Monte Carlo, genetic type search algorithms, island particle models, Gibbs cloning search techniques, interacting simulated annealing algorithms, importance sampling methods, branching and splitting particle algorithms, rare event simulations, quantum and diffusion Monte Carlo models, adaptive population Monte Carlo sampling models, Ensemble Kalman filters and interacting Kalman filters.

Since computations are nowadays much more affordable, the aforementioned particle methods have become revolutionary for solving complex estimation and optimization problems arising in engineering, risk analysis, Bayesian statistics and information sciences. The books [46], [47], [53],[65] provide a rather complete review on these application domains.

These topics have constituted some of the main research axes of several of the ASTRAL team members since the beginning of the 1990s. To the best of our knowledge, the first rigorous study on particle filters and the convergence of genetic algorithms as the size of the population tends to infinity seems to be the article [55], published in 1996 in the journal Markov Processes and Related Fields. This paper has opened an avenue of research questions in stochastic analysis and particle methods applications. The uniform convergence of particle filters and ensemble Kalman filters with respect to the time horizon was first seen in [50, 49, 52] and in the more recent article [54]. The first use of particle algorithms and Approximated Bayesian Computation type methodologies in nonlinear filtering seems to have started in [51]. Last but not least, the development of sequential Monte Carlo methodology in statistics was introduced in the seminal article [48].

Despite some recent advances, the mathematical foundation and the design and the numerical analysis of stochastic particle methods is still a significant challenge. For instance, particle filter technology is often combined with Metropolis-Hastings type techniques, or with Expectation Maximization type algorithms. The resulting algorithms are intended to solve high dimensional hidden Markov chain problems with fixed parameters. In this context (despite some recent attempts) the refined convergence analysis of the resulting particle algorithms, including exponential concentration estimates, remains to be developed.

Last but not least, the expectations of their performances are constantly rising in a variety of application domains. These particle methodologies are now expected to deal with increasingly sophisticated models in high dimensions, whilst also allowing for the variables to evolve at different scales. *The overarching aim of this aspect of the programme is to make a breakthrough in both the mathematical analysis and the numerical simulation of stochastic and interacting particle algorithms.*

Today, partly because of the emergence of new mean field simulation methodologies and partly because of the importance of new and challenging high-dimensional problems arising in statistical machine learning, engineering sciences and molecular chemistry, we are observing the following trends:

• A need to better calibrate their performance with respect to the size of the systems and other tuning parameters, including cooling decay rates, local random search strategies, interacting and adaptive search criteria, and population size parameters. One of the main and central objectives is to obtain uniform and non asymptotic precision estimates with respect to the time parameter. These types of uniform estimates need to be developed, supporting industrial goals of enhanced design and confidence of algorithms, risk reduction and improved safety.

• A need for new stochastic and adaptive particle methods for solving complex estimation models. Such models arise in a range of application areas including forecasting, data assimilation, financial risk management and analysis of critical events. This subject is also crucial in environmental studies and in the reliability analysis of engineering automated systems. The complexity of realistic stochastic models in advanced risk analysis requires the use of sophisticated and powerful stochastic particle models. These models go far beyond Gaussian models, taking into account abrupt random changes, as well as non nonlinear dynamics in high dimensional state spaces.

• A need to find new mathematical tools to analyze the stability and robustness properties of sophisticated, nonlinear stochastic models involving space-time interaction mechanisms. Most of the theory on the stability of Markov chains is based on the analysis of the regularity properties of linear integral semigroups. To handle these questions, the interface between the theory of nonlinear dynamical systems and the analysis of measure valued processes needs to be further developed.

From a purely probabilistic point of view, the fundamental and the theoretical aspects of our research projects are essentially based on the stochastic analysis of the following three classes of interacting stochastic processes: *Spatial branching processes and mean-field type interacting particle systems, reinforced and self-interacting processes, and finally random tree based search/smoothing learning processes.*

The first class of particle models includes interacting jump-diffusions, discrete generation models, particle ensemble Kalman filters and evolutionary algorithms. This class of models refers to mean field type interaction processes with respect to the occupation measure of the population. For instance genetic-type branching-selection algorithms are built on the following paradigm: when exploring a state space with many particles, we duplicate better fitted individuals while particles with poor fitness die. The selection is made by choosing randomly better fitted individuals in the population. Our final aim is to develop a complete mean-field particle theory combining the stability properties of the limiting processes as the size of the system tends to infinity with the performance analysis of these particle sampling tools.

The second class of particle models refers to mean field type interaction processes with respect to the occupation measure of the past visited sites. This type of reinforcement is observed frequently in nature and society, where "beneficial" interactions with the past history tend to be repeated. Self interaction gives the opportunity to build new stochastic search algorithms with the ability to, in a sense, re-initialize their exploration from the past, re-starting from some better fitted previously visited initial value. In this context, we plan to explore the theoretical foundations and the numerical analysis of continuous time or discrete generation self-organized systems by combining spatial and temporal mean field interaction mechanisms.

The last generation of stochastic random tree models is concerned with biology-inspired algorithms on paths and excursions spaces. These genealogical adaptive search algorithms coincide with genetic type particle models in excursion spaces. They have been successfully applied in generating the excursion distributions of Markov processes evolving in critical and rare event regimes, as well as in path estimation and related smoothing problems arising in advanced signal processing. The complete mathematical analysis of these random tree models, including their long time behavior, their propagation of chaos properties, as well as their combinatorial structures are far from complete.

Our research agenda on stochastic learning is developed around the applied mathematical axis as well as the numerical perspective, including concrete industrial transfers with a special focus on Naval Group. From the theoretical side, mid-term objectives are centered around non asymptotic performance analysis and long time behavior of Monte Carlo methods and stochastic learning algorithms. We also plan to further develop the links with Bayesian statistical learning methodologies and artificial intelligence techniques, including the analysis of genetic programming discussed in section 3.1. We also have several long term projects. The first one is to develop new particle type methodologies to solve high dimensional data assimilation problems arising in forecasting and fluid mechanics, as well as in statistical machine learning. We also plan to design stochastic filtering-type algorithms to solve partially observed control problems such as those discussed in section 3.3.

Project-team positioning: In the last three decades, the use of Feynman-Kac type particle models has been developed in variety of scientific disciplines, including in molecular chemistry, risk analysis, biology, signal processing, Bayesian inference and data assimilation.

The design and the mathematical analysis of Feynman-Kac particle methodologies has been one of the main research topics of P. Del Moral since the late 1990's [55, 51, 50], see also the books [52, 46, 47, 53] and references therein. These mean field particle sampling techniques encapsulate particle filters, sequential Monte Carlo methods, spatial branching and evolutionary algorithms, Fleming-Viot genetic type particles methods arising in the computation of quasi-invariant measures and simulation of non absorbed processes, as well as diffusion Monte Carlo methods arising in numerical physics and molecular chemistry. The term "particle filters" was first coined in the article [55] published in 1996 in reference to branching and mean field interacting particle methods used in fluid mechanics since the beginning of the 1960s. This article presents the first rigorous analysis of these mean field type particle algorithms.

The INRIA project teams applying the particle methodology developed by ASTRAL include the INRIA project team SIMSMART (rare event simulation as well as particle filters) and the INRIA project team

Matherials (applications in molecular chemistry). The project team ASTRAL also has several collaborative research projects with these, teams as well as with researchers from international universities working in this subject, including Oxford, Cambridge, New South Wales Sydney, UTS, Bath, Warwick and Singapore Universities.

3.3 Decision and stochastic control

Permanent researchers: P. Del Moral, F. Dufour, A. Genadot, E. Horton, D. Laneuville, O. Marceau, A. Nègre, J. Saracco, H. Zhang.

Part of this research project is devoted to the analysis of stochastic decision models. Many real applications in dynamic optimization can be, roughly speaking, described in the following way: a certain system evolves randomly under the control of a sequence of actions with the objective to optimize a performance function. Stochastic decision processes have been introduced in the literature to model such situations and it is undoubtedly their generic capacity to model real life applications that leads to and continues to contribute to their success in many fields such as engineering, medicine and finance.

In this project we will focus on specific families of models that can be identified according to the following elements: the nature of the time variable (discrete or continuous), the type of dynamics (piecewise deterministic trajectories) and the numbers of decision makers. For one player, the system will be called a *stochastic control process* and for the case of several decision-makers, the name *(stochastic) game* will be used. For ease of understanding, we now provide an informal description of the classes of stochastic processes we are interested in, according to the nature of the time variable.

Discrete-time models. In this framework, the basic model can be described by a state space where the system evolves, an action space, a stochastic kernel governing the dynamic and, depending on the state and action variables, a one-step cost (reward) function. The distribution of the controlled stochastic process is defined through the control policy which is then selected in order to optimize the objective function. This is a very general model for dynamic optimization in discrete-time, which also goes by the name of *stochastic dynamic programming*. For references, the interested reader may consult the following books [41, 43, 56, 57, 59, 60, 61, 62, 67, 66, 69] and the references therein (this list of references is, of course, not exhaustive).

Continuous-time models. Most of the continuous-time stochastic processes consist of a combination of the following three different ingredients: stochastic jumps, diffusion and deterministic motions. In this project, we will focus on **non-diffusive models**, in other words, stochastic models for which the randomness appears only at fixed or random times, *i.e.* those combining deterministic motions and random jumps. These stochastic processes are the so-called piecewise deterministic Markov processes (PDMPs) [42, 44, 45, 58, 64, 63, 68]. This family of models plays a central role in applied probability because it forms the bulk of models in many research fields such as, *e.g.* operational research, management science and economy and covers an enormous variety of applications.

These models can be framed in several different forms of generality, depending on their mathematical properties such as the type of performance criterion, full or incomplete state information, with or without constraints, adaptative or not, but more importantly, the nature of the boundary of the state space, the type of dynamic between two jumps and on the number of decision-makers. These last three characteristics make the analysis of the controlled process much more involved.

Part of this project will cover both theoretical and numerical aspects of stochastic optimal control. It is clear that stochastic problems and control games have been extensively studied in the literature. Nevertheless, important challenges remain to be addressed. From the theoretical side, there are still many technical issues that are, for the moment, still unanswered or at most have received partial answers. This is precisely what makes them difficult and requires either the creative transposition of pre-existing methodologies or the development of new approaches. It is interesting to note that one of the feature of these theoretical problems is that they are closely related to practical issues. Solving such problems not only gives rise to challenging mathematical questions, but also allow a better insight into the structure and properties of real practical problems. Theory for applications will be for us the thrust that will guide us in this project. From the numerical perspective, solving a stochastic decision model remains a critical issue. Indeed, except for very few specific models, the determination of an optimal policy and the associated value function is an extremely difficult problem to tackle. The development of computational and numerical methods to get quasi-optimal solutions is, therefore, of crucial importance to demonstrate the practical interest of stochastic decision model as a powerful modeling tool. During the International Conference on Dynamic Programming and Its Applications held at the University of British Columbia, Canada in April 1977, Karl Hinderer, a pioneer in the field of stochastic dynamic programming emphasized that "*whether or not our field will have a lasting impact on science beyond academic circles depends heavily on the success of implemented applications*". We believe that this statement is still in force some forty years later.

The objective of this project is to address these important challenges. They are mainly related to models with general state/action spaces and with continuous time variables covering a large field of applications. Here is a list of topics we would like to study: games, constrained control problems, non additive types of criteria, numerical and computational challenges, analysis of partially observed/known stochastic decision processes. This list is not necessarily exhaustive and may of course evolve over time.

Our research agenda on optimal stochastic control is developed around the applied mathematical axis as well as the numerical perspective, including concrete industrial transfers with a special focus on Naval Group. Our mid-term objectives will focus on the following themes described above: properties of control policies in continuous-time control problems, non additive types of criteria, numerical and computational challenges. Our long-term objectives will focus on the analysis of partially observed/known stochastic control problems, constrained control problems and games.

Project-team positioning: There exists a large national/international community working on PDMPs and MDPs both on the theoretical, numerical and practical aspects. One may cite A. Almudevar (University of Rochester, USA), E. Altman (INRIA Team NEO, France), K. Avrachenkov (INRIA Team NEO, France), N. Bauerle (Karlsruhe University, Germany), D. Bertsekas (Massachusetts Institute of Technology, USA), O. Costa (Sao Paulo University, Brazil), M. Davis (Imperial College London, England), E. Feinberg (Stony Brook University, USA), D. Goreac (Université Paris-Est Marne-la-Vallée, France), X. Guo (Zhongshan University, China), O. Hernandez-Lerma (National Polytechnic Institute, Mexico), S. Marcus (University of Maryland, USA), T. Prieto-Rumeau (Facultad de Ciencias, UNED, Spain), A. Piunovskiy (University of Liverpool, England), U. Rieder (Universität Ulm, Germany), J. Tsitsiklis (Massachusetts Institute of Technology, USA), B. Van Roy (Stanford University, USA), O. Vega-Amaya (Universidad de Sonora, Mexico), Y. Zhang (University of Liverpool, England) to name just a few. Many of the colleagues cited above are at the head of research groups which have not been described in detail due to space limitation and so, this list is far from being exhaustive.

To some extent, our team is in competition wit the colleagues and teams mentioned above. We emphasize that there exists a long standing collaboration between our group and O. Costa (Sao Paulo University, Brazil) since 1998. In the last 10 years, we have established very fruitful collaborations with T. Prieto-Rumeau (Facultad de Ciencias, UNED, Spain) and A. Piunovskiy (University of Liverpool, England).

Inside INRIA, the team NEO and in particular E. Altman and K. Avrachenkov work on discrete-time MDPs but they are mainly focused on the case of countable (finite) state/action spaces MDPs. From this point of view, our results on this theme may appear complementary to theirs.

4 Application domains

It is important to point out that (for the time being) only a sub-group of the academic part of the team collaborates with Naval Group. Initially the topics of interest for Naval Group was focused on filtering and control problems. The academic members of this sub-group are P. Del Moral, F. Dufour, A. Genadot, E. Horton and H. Zhang. It is also important to emphasize that Naval Group is undoubtedly our privileged industrial partner. This collaboration is described in section 4.1. For reasons of confidentiality, this section is not very detailed, in particular it does not mention the timetable and does not detail the technical solutions that will be considered. Our aim in the short term is to integrate the remaining academic team members into the group to work on the themes of interest to NG. A seminar was organized for this purpose in August 2020. The academic members of the team who are not involved in collaboration with NG (M. Chavent, P. Legrand and J. Saracco) have their own industrial collaborations that are described in section 4.2.

4.1 Naval Group research activities

Permanent researchers: P. Del Moral, F. Dufour, A. Genadot, E. Horton, D. Laneuville, O. Marceau, A. Nègre and H. Zhang.
An important line of research of the team is submarine passive target tracking. This is a very complicated practical problem that combines both filtering and stochastic control topics. In the context of passive underwater acoustic warfare, let us consider a submarine, called the observer, equipped with passive sonars collecting noisy bearing-only measurements of the target(s). The trajectory of the observer has to be controlled in order to satisfy some given mission objectives. These can be, for example, finding the best trajectory to optimize the state estimation (position and velocity) of the targets, maximize the different targets' detection range and/or minimize its own acoustic indiscretion with respect to these targets, and reaching a way-point without being detected. Let us now describe in more detail some of the topics we intend to work on.

In the case of passive tracking problems, one of the main issues is that the observer must manoeuvre in order to generate observability. It turns out that these manoeuvres are actually necessary but not sufficient to guarantee that the problem becomes observable. In fact, a significant body of the literature pertains to attempting to understand whether this type of problem is solvable. Despite this observability analysis, the following practical questions, which we would like to address in this project, remain challenging: What kind of trajectory should the observer follow to optimize the estimation of the target's motion? What is the accuracy of that estimate? How to deal with a multitarget environment? How to take into account some physical constraints related to the sonar?

Another aspect of target tracking is to take into account both the uncertainties on the target's measurement and also the signal attenuation due to acoustic propagation. To the best of our knowledge, there are few works focusing on the computation of optimal trajectories of underwater vehicles based on signal attenuation. In this context, we would like address the problem of optimizing the trajectory of the observer to maximize the detection of the acoustic signals issued by the targets. Conversely, given that the targets are also equipped with sonars, how can one optimize the trajectory of the observer itself to keep its own acoustic indiscretion as low as possible with respect to those targets.

It must be emphasized that a human operator can find a suitable trajectory for either of these objectives in the context of a single target. However, if both criteria and/or several targets are taken into account simultaneously, it is hardly possible for a human operator to find such trajectories. From an operational point of view, these questions are therefore of great importance.

Such practical problems are strongly connected to the mathematical topics described in sections 3.2 and 3.3. For example it is clearly related to partially observed stochastic control problems. The algorithmic solutions that we will develop in the framework of submarine passive target tracking will be evaluated on the basis of case studies proposed by Naval Group. Our short-term aim is to obtain explicit results and to develop efficient algorithms to solve the various problems described above.

4.2 Other collaborations

Permanent researchers: M. Chavent, P. Legrand and J. Saracco.
For several years, the team has also had strong collaborations with INRAE which is the French National Research Institute for Agriculture, Food and Environment. More precisely, consumer satisfaction when eating beef is a complex response based on subjective and emotional assessments. Safety and health are very important in addition to taste and convenience but many other parameters are also extremely important for breeders. Many models were recently developed in order to predict each quality trait and to

evaluate the possible trade-off that could be accepted in order to satisfy all the operators of the beef chain at the same time. However, in none of these quality prediction systems are issues of joint management of the different expectations addressed. Thus, it is vital to develop a model that integrates the sensory quality of meat but also its nutritional and environmental quality, which are expectations clearly expressed by consumers. Our team are currently developing statistical models and machine learning tools in order to simultaneously manage and optimize the different sets of expectations. Combining dimension reduction methodologies, nonparametric quantiles estimation and "Pareto front" approaches could provide an interesting way to address this complex problem. These different aspects are currently in progress.

The team is currently initiating scientific collaboration with the Advanced Data Analytics Group of Sartorius Corporate Research which is an international pharmaceutical and laboratory equipment supplier, covering the segments of Bioprocess Solutions and Lab Products & Services. The current work concerns the development of a partial least squares (PLS) inspired method in the context of multiblocks of covariates (corresponding to different technologies and/or different sampling techniques and statistical procedures) and high dimensional datasets (with the sample size n much smaller than the number of variables in the different blocks). The proposed method allows variable selection in the X and in the Y components thanks to interpretable parameters associated with the soft-thresholding of the empirical correlation matrices (between the X 's blocks and the Y block) decomposed using singular values decomposition (SVD). In addition, the method is able to handle specific missing values (i.e. "missing samples" in some covariate blocks). The suggested `ddsPLS + Koh Lanta` methodology is computationally fast. Some technical and/or theoretical work on this methodology must be naturally pursued in order to further refine this approach. Moreover, another aspect of the future research with Sartorius consists of associating the structures of datasets with the real biological dynamics described, until now, by differential equations and for which the most advanced solutions do not merge with both high dimensional multiblock analysis and these differential equations. Combining these two approaches in a unified framework will certainly have many applications in industry and especially in the biopharmaceutical production.

Within the framework of the GIS ALBATROS, the team has initiated a scientific collaboration with IMS and THALES. The first topic is focused on the measurement of the cognitive load of a pilot through the development of methods for measuring the regularity of biological signals (Hölderian regularity, Detrended Fluctuation Analysis, etc.). The second topic is dedicated to the development of classification techniques of vessels. The different methods we proposed are based on deep learning, evolutionary algorithms and signal processing techniques and are compared to the approaches in the literature.

5 Highlights of the year

Pierre Del Moral has been a Distinguished Visiting Professors at Heilbronn Institute for Mathematical Research. During his visit he gave a seminar: Stability of positive semigroups and their mean field interpretations

6 New results

6.1 Statistical learning

6.1.1 Multivariate Analysis of Mixed Data: The R Package PCAmixdata

Mixed data type arise when observations are described by a mixture of numerical and categorical variables. The R package PCAmixdata extends standard multivariate analysis methods to incorporate this type of data. The key techniques included in the package are PCAmix (PCA of a mixture of numerical and categorical variables), PCArot (rotation in PCAmix) and MFAmix (multiple factor analysis with mixed data within a dataset). In this paper [11] a synthetic presentation of the three algorithms will be provided and the three main procedures will be illustrated on real data composed of four datasets caracterizing conditions of life of cities of Gironde, a south-west region of France.

Participants: Marie Chavent (ASTRAL), Vanessa Kuentz-Simonet, Amaury Labenne, Jérôme Saracco (ASTRAL).

6.1.2 Advanced topics in Sliced Inverse Regression

Since its introduction in the early 90's, the Sliced Inverse Regression (SIR) methodology has evolved adapting to increasingly complex data sets in contexts combining linear dimension reduction with non linear regression. The assumption of dependence of the response variable with respect to only a few linear combinations of the covariates makes it appealing for many computational and real data application aspects. This work [16] proposes an overview of the most active research directions in SIR modeling from multivariate regression models to regularization and variable selection.

Participants: Stéphane Girard, Hadrien Lorenzo (ASTRAL), Jérôme Saracco (ASTRAL).

6.2 Stochastic control and games

6.2.1 On the equivalence of the integral and differential Bellman equations in impulse control problems

When solving optimal impulse control problems, one can use the dynamic programming approach in two different ways: at each time moment, one can make the decision whether to apply a particular type of impulse, leading to the instantaneous change of the state, or apply no impulses at all; or, otherwise, one can plan an impulse after a certain interval, so that the length of that interval is to be optimised along with the type of that impulse. The first method leads to the differential Bellman equation, while the second method leads to the integral Bellman equation. In this paper [13] we prove the equivalence of those Bellman equations in many specific models. Those include abstract dynamical systems, controlled ordinary differential equations, piece-wise deterministic Markov processes and continuous-time Markov decision processes.

Participants: François Dufour (ASTRAL), Alexei Piunovskiy, Alexander Plakhov.

6.2.2 Maximizing the probability of visiting a set infinitely often for a countable state space Markov decision process

In [14], we consider a Markov decision process with countable state space and Borel action space. We are interested in maximizing the probability that the controlled Markov chain visits some subset of the state space infinitely often. We provide sufficient conditions, based on continuity and compactness requirements, together with a stability condition on a parametrized family of auxiliary control models, which imply the existence of an optimal policy that is deterministic and stationary. We compare our hypotheses with those existing in the literature.

Participants: François Dufour (ASTRAL), Tomas Prieto-Rumeau.

6.2.3 Stationary Markov Nash equilibria for nonzero-sum constrained ARAT Markov games

In [14], we consider a nonzero-sum Markov game on an abstract measurable state space with compact metric action spaces. The goal of each player is to maximize his respective discounted payoff function under the condition that some constraints on a discounted payoff are satisfied. We are interested in the existence of a Nash or noncooperative equilibrium. Under suitable conditions, which include absolute continuity of the transitions with respect to some reference probability measure, additivity of the payoffs and the transition probabilities (ARAT condition), and continuity in action of the payoff functions and the density function of the transitions of the system, we establish the existence of a constrained stationary Markov Nash equilibrium, that is, the existence of stationary Markov strategies for each of the players yielding an optimal profile within the class of all history-dependent profiles.

Participants: François Dufour (ASTRAL), Tomas Prieto-Rumeau.

6.2.4 Optimal Control of Piecewise Deterministic Markov Processes

In this Book's chapter [33] we study the infinite-horizon continuous-time optimal control problem of piece- wise deterministic Markov processes (PDMPs) with the control acting continuously on the jump intensity λ and on the transition measure Q of the process. Two optimality criteria are considered, the discounted cost case and the long run average cost case. We provide conditions for the existence of a solution to an integro-differential optimality equality, the so called Hamilton-Jacobi-Bellman (HJB) equation, for the discounted cost case, and a solution to an HJB inequality for the long run average cost case, as well as conditions for the existence of a deterministic stationary optimal policy. From the results for the discounted cost case and under some continuity and compactness hypothesis on the parameters and non-explosive assumptions for the process, we derive the conditions for the long run average cost case by employing the so-called vanishing discount approach.

Participants: Oswaldo Do Valle Costa, François Dufour (ASTRAL).

6.3 Stochastic learning

6.3.1 Log-normalization constant estimation using the ensemble Kalman–Bucy filter with application to high-dimensional models

In this article [12], we consider the estimation of the log-normalization constant associated to a class of continuous-time filtering models. In particular, we consider ensemble Kalman-Bucy filter based estimates based upon several nonlinear Kalman-Bucy diffusions. Based upon new conditional bias results for the mean of the afore-mentioned methods, we analyze the empirical log-scale normalization constants in terms of their Ln-errors and conditional bias. Finally, we use these results for online static parameter estimation for above filtering models and implement the methodology for both linear and nonlinear models.

Participants: Dan Crisan, Pierre del Moral (ASTRAL), Ajay Jasra, Hamza Ruzayqat.

6.3.2 Asymptotic moments of spatial branching processes

In this paper [17], it is assumed that $X = (X_t, t \geq 0)$ is either a superprocess or a branching Markov process on a general space E, with non-local branching mechanism and probabilities P_{δ_x} , when issued from a unit mass at $x \in E$. For a general setting in which the first moment semigroup of X displays a Perron-Frobenius type behaviour, we show that for $k \geq 2$ and any positive bounded measurable function f on E, $\lim_{t\to\infty} g(t)E_{\delta_x}[(f, X_t)^k] = C_k(x, f)$ where the constant $C_k(x, f)$ can be identified in terms of the principal right eigenfunction and left eigen-measure and $g(t)$ is an appropriate determinisitic normalisation, which can be identified explicitly as either polynomial in t or exponential in t, depending on whether X is a critical, supercritical or subcritical process. The method we employ is extremely robust and we are able to extract similarly precise results that additionally give us the moment growth with time of $\int_0^t (g, X_t) ds$, for bounded measurable g on E.

Authors: Isaac Gonzalez, Emma Horton (ASTRAL), Andreas E Kyprianou

6.3.3 Yaglom limit for critical nonlocal branching markov processes

In [19], we consider the classical Yaglom limit theorem for the Neutron Branching Process (NBP) in the setting that the mean semigroup is critical, i.e. its leading eigenvalue is zero. We show that the law of the process conditioned on survival is asymptotically equivalent to an exponential distribution. As part of the proof, we also show that the probability of survival decays inversely proportionally to time. Although Yaglom limit theorems have recently been handled in the setting of spatial branching processes and superprocesses, as well as in the setting of isotropic homogeneous Neutron Branching Processes, our approach and the main novelty of this work is based around a precise result for the scaled asymptotics for the k-th martingale moments of the NBP (rather than the Yaglom limit itself). Our proof of the asymptotic martingale moments turns out to offer a general approach to asymptotic martingale moments of critical branching Markov processes with a non-local branching mechanism. Indeed this is the context in which we give both our moment proofs and the Yaglom limit.

Authors: Simon C Harris, Emma Horton (ASTRAL), Andreas E Kyprianou, Minmin Wang.

6.3.4 Strong laws of large numbers for a growth-fragmentation process with bounded cell sizes

Growth-fragmentation processes model systems of cells that grow continuously over time and then fragment into smaller pieces. Typically, on average, the number of cells in the system exhibits asynchronous exponential growth and, upon compensating for this, the distribution of cell sizes converges to an asymptotic profile. However, the long-term stochastic behaviour of the system is more delicate, and its almost sure asymptotics have been so far largely unexplored. In this article [20] , we study a growth-fragmentation process whose cell sizes are bounded above, and prove the existence of regimes with differing almost sure long-term behaviour.

Participants: Emma Horton (ASTRAL), A. R. Watson.

6.3.5 A note on Riccati matrix difference equations

Discrete algebraic Riccati equations and their fixed points are well understood and arise in a variety of applications, however, the time-varying equations have not yet been fully explored in the literature. In this article [21] we provide a self-contained study of discrete time Riccati matrix difference equations. In particular, we provide a novel Riccati semigroup duality formula and a new Floquet-type representation for these equations. Due to the aperiodicity of the underlying flow of the solution matrix, conventional Floquet theory does not apply in this setting and thus further analysis is required. We illustrate the impact of these formulae with an explicit description of the solution of time-varying Riccati difference equations and its fundamental-type solution in terms of the fixed point of the equation and an invertible linear matrix map, as well as uniform upper and lower bounds on the Riccati maps. These are the first results of this type for time varying Riccati matrix difference equations.

Participants: Pierre Del Moral (ASTRAL), Emma Horton (ASTRAL).

6.3.6 A theoretical analysis of one-dimensional discrete generation ensemble Kalman particle filters

Despite the widespread usage of discrete generation Ensemble Kalman particle filtering methodology to solve nonlinear and high dimensional filtering and inverse problems, little is known about their mathematical foundations. As genetic-type particle filters (a.k.a. sequential Monte Carlo), this ensemble-type methodology can also be interpreted as mean-field particle approximations of the Kalman-Bucy filtering equation. In contrast with conventional mean-field type interacting particle methods equipped with a globally Lipschitz interacting drift-type function, Ensemble Kalman filters depend on a nonlinear and quadratic-type interaction function defined in terms of the sample covariance of the particles. Most of the literature in applied mathematics and computer science on these sophisticated interacting particle methods amounts to designing different classes of useable observer-type particle methods. These methods are based on a variety of inconsistent but judicious ensemble auxiliary transformations or include additional inflation/localisationtype algorithmic innovations, in order to avoid the inherent time-degeneracy of an insufficient particle ensemble size when solving a filtering problem with an unstable signal. To the best of our knowledge, the first and the only rigorous mathematical analysis of these sophisticated discrete generation particle filters is developed in the pioneering articles by Le Gland-Monbet-Tran and by Mandel-Cobb-Beezley, which were published in the early 2010s. Nevertheless, besides the fact that these studies prove the asymptotic consistency of the Ensemble Kalman filter, they provide exceedingly pessimistic meanerror estimates that grow exponentially fast with respect to the time horizon, even for linear Gaussian filtering problems with stable one dimensional signals. In the present article [22] we develop a novel self-contained and complete stochastic perturbation analysis of the fluctuations, the stability, and the long-time performance of these discrete generation ensemble Kalman particle filters, including time-uniform and non-asymptotic mean-error estimates that apply to possibly unstable signals. To the best of our knowledge, these are the first results of this type in the literature on discrete generation particle filters, including the class of genetic-type particle filters and discrete generation ensemble Kalman filters. The stochastic Riccati difference equations considered in this work are also of interest in their own right, as a prototype of a new class of stochastic rational difference equation.

Participants: Pierre Del Moral (ASTRAL), Emma Horton (ASTRAL).

6.3.7 On the Stability of Positive Semigroups

In [23], the stability and contraction properties of positive integral semigroups on locally compact Polish spaces are investigated. We provide a novel analysis based on an extension of V-norm, Dobrushin-type, contraction techniques on functionally weighted Banach spaces for Markov operators. These are applied to a general class of positive and possibly time-inhomogeneous bounded integral semigroups and their normalised versions. Under mild regularity conditions, the Lipschitz-type contraction analysis presented in this article simplifies and extends several exponential estimates developed in the literature. The spectraltype theorems that we develop can also be seen as an extension of Perron-Frobenius and Krein-Rutman theorems for positive operators to time-varying positive semigroups. We review and illustrate in detail the impact of these results in the context of positive semigroups arising in transport theory, physics, mathematical biology and advanced signal processing.

Participants: Pierre Del Moral (ASTRAL), Emma Horton (ASTRAL), Ajay Jasra.

6.3.8 A Review of Neural Network-Based Emulation of Guitar Amplifiers

In [24], Review of Neural Network-Based Emulation of Guitar Amplifiers is proposed. Vacuum tube amplifiers present sonic characteristics frequently coveted by musicians, that are often due to the distinct nonlinearities of their circuits, and accurately modelling such effects can be a challenging task. A recent rise in machine learning methods has lead to the ubiquity of neural networks in all fields of study including virtual analog modelling. This has lead to the appearance of a variety of architectures tailored to this task. This work aims to provide an overview of the current state of the research in neural emulation of analog distortion circuits by first presenting preceding methods in the field and then focusing on a complete review of the deep learning landscape that has appeared in recent years, detailing each subclass of available architectures. This is done in order to bring to light future possible avenues of work in this field.

Participants: Tara Vanhatalo (ASTRAL), Pierrick Legrand (ASTRAL), Myriam Desainte-Catherine, Pierre Hanna, Antoine Brusco, Guillaume Pille, Yann Bayle.

7 Bilateral contracts and grants with industry

7.1 Bilateral contracts with industry

Naval Group

Participants: Pierre Del Moral, François Dufour, Alexandre Genadot, Emma Horton, Dann Laneuville, Olivier Marceau, Adrien Nègre, Raymond Zhang.

In the application domain, an important research focus of the team is the tracking of passive underwater targets in the context of passive underwater acoustic warfare. This is a very complicated practical problem that combines both filtering and stochastic control issues. This research topic is addressed in collaboration with Naval Group. We refer the reader to the section 4.1 for a more detailed description of this theme.

Thales AVS

Participants: Bastien Berthelot, Pierrick Legrand.

The collaboration is centered around some contributions to the estimation of the Hurst coefficient and his application on biosignals in the domain of crew monitoring.

Case Law Analytics

Participants: Pierrick Legrand.

Pierrick Legrand is a consultant for the startup Case Law Analytics. The object of the consulting is confidential.

Sartorius

Participants: Hadrien Lorenzo, Jérôme Saracco.

The team is currently initiating a scientific collaboration with the Advanced Data Analytics Group of Sartorius Corporate Research which is an international pharmaceutical and laboratory equipment supplier, covering the segments of Bioprocess Solutions and Lab Products and Services. The current work concerns the development of a PLS (Partial Least Squares) inspired method in the context of multiblock of covariates (corresponding to different technologies and/or different sampling, statistical natures. . .) and high dimensional datasets (with the sample size n much smaller than the number of variables in the different blocks). The proposed method, called ddsPLS for data-driven sparse PLS, allows variable selection in the X and in the Y parts thanks to interpretable parameters associate with the soft-thresolding of the empirical correlation matrices (between the X's blocks and the Y block) decomposed in SVD (Singular Values Decomposition) ways. In addition a methodology to handle specific missing values (i.e. missing samples in some covariate blocks) is also under investigation.

Safran Aircraft Engines

Participants: Marie Chavent, Jérôme Lacaille, Alex Mourer, Madalina Olteanou.

The collaboration is centered around an applied mathematics thesis defining a formalism and a methodology for processing and interpretation by the importance of variables (from measurements and calculated indicators) in the case of unsupervised problems. This methodology is accompanied by code programming and a demonstration on an example data set from Safran Aircraft Engines.

7.2 Bilateral Grants with Industry

Orosys

Participants: Tara Vanhatalo, Pierrick Legrand.

Within the framework of Tara Vanhatalo's Cifre PhD thesis on the stochastic modelingof guitar amplifiers, a strong collaboration was established between the company Orosys and the ASTRAL team.

8 Partnerships and cooperations

8.1 International initiatives

Participants: Emma Horton.

8.1.1 Inria associate team not involved in an IIL or an international program

MAGO

Title: Modelling and analysis for growth-fragmentation processes

Duration: 2022 ->

Coordinator: Alexander Watson (alexander.watson@ucl.ac.uk)

Partners:

- University College London London (Royaume-Uni)

Inria contact: Emma Horton

Summary: Growth-fragmentation (GF) refers to a collection of mathematical models in which objects – classically, biological cells – slowly gather mass over time, and fragment suddenly into multiple, smaller offspring. These models may be used to represent a range of biological and chemical processes, in which an individual reproduces by fission into two or more new individuals, such as the the evolution of plasmids in bacteria populations and protein polymerisation.

As such, in recent years, there has been a growing interest in probabilistic models, in the form of GF processes, in order to gain a better understanding of these real-world processes. However, the complexity of the real-world models and thus of the GF processes makes them highly intractable, requiring the development of sophisticated probabilistic and statistical techniques to analyse these models.

In particular, it is crucial to understand the behaviour of GF models in order to develop algorithms to simulate real-world processes, estimate quantities such as the growth rate of the system and the steady state behaviour, and for the purpose of parameter inference, to allow scientists to gain a better understanding of the behaviour of these complex systems.

In this proposal, we aim to develop a suite of GF processes that capture the many complexities and nuances of a range of fundamental biological and chemical processes. We will then analyse the macroscopic and microscopic behaviour of these stochastic systems, developing new simulation techniques, methods for parameter inference and goodness of fit tests, for use with real data.

8.2 International research visitors

8.2.1 Visits of international scientists

Other international visits to the team

Alexander Cox

Status Professor

Institution of origin: University of Bath

Country: UK

Dates: 12/10/22 - 14/10/22

Context of the visit: To discuss new projects arising from an existing collaboration on interacting particle systems.

Mobility program/type of mobility: Research stay

Alastair Crossley

Status Master student

Institution of origin: University of Bath

Country: UK

Dates: 19/09/22 - 26/09/22

Context of the visit: Summer project on sensitivity analysis of branching processes.

Mobility program/type of mobility: Research stay as part of summer internship.

Sophie Hautphenne

Status Senior Lecturer

Institution of origin: University of Melbourne

Country: Australia

Dates: 31/07/22 - 08/08/22

Context of the visit: Collaboration on parameter estimation for branching processes.

Mobility program/type of mobility: Research stay as part of PHC FASIC Chercheur programme.

Oscar Key

Status PhD

Institution of origin: UCL

Country: UK

Dates: 12/12/22 - 14/12/22

Context of the visit: To develop simulation techniques for growth-fragmentation processes.

Mobility program/type of mobility: Research stay as part of Inria Associate Team MAGO.

Bastien Mallein

Status Maître de conférence

Institution of origin: University of Paris XIII

Country: France

Dates: 26/10/22 - 28/10/22

Context of the visit: To discuss new projects concerning the long-term behaviour of branching processes.

Mobility program/type of mobility: Research stay

Pierre Nyquist

Status Associate Professor

Institution of origin: KTH

Country: Sweden

Dates: 12/10/22 - 14/10/22

Context of the visit: To discuss new projects arising from an existing collaboration on interacting particle systems.

Mobility program/type of mobility: Research stay

Ellen Powell

Status Associate Professor

Institution of origin: University of Durham

Country: UK

Dates: 25/07/22 - 29/07/22, 4/10/22 - 7/10/22

Context of the visit: Collaboration on genealogies of branching processes.

Mobility program/type of mobility: Research stay as part of PHC Alliance programme.

Alexander Watson

Status Associate Professor

Institution of origin: UCL

Country: UK

Dates: 12/12/22 - 14/12/22

Context of the visit: To develop simulation techniques for growth-fragmentation processes.

Mobility program/type of mobility: Research stay as part of Inria Associate Team MAGO.

8.2.2 Visits to international teams

Research stays abroad

Participants: Emma Horton.

Visited institution: KTH

Country: Sweden

Dates: 07/03/22 - 09/03/22

Context of the visit: To discuss new projects on large deviation effects arising in interacting particle systems. Seminar.

Mobility program/type of mobility: Research stay and seminar.

Visited institution: University of Bath

Country: UK

Dates: 11/04/22 - 15/04/22

Context of the visit: Workshop on Polya urns, stochastic approximation and quasi-stationary distributions.

Mobility program/type of mobility: Workshop

Visited institution: Université de Montréal

Country: Canada

Dates: 02/05/22 - 12/05/22

Context of the visit: Workshop on branching structures.

Mobility program/type of mobility: Workshop.

Visited institution: University of Bath

Country: UK

Dates: 13/06/22 - 17/06/22

Context of the visit: Workshop to develop new collaborations with industry.

Mobility program/type of mobility: Workshop.

Visited institution: University of Bath

Country: UK

Dates: 15/08/22 - 09/09/22

Context of the visit: Supervision of master student on sensitivity analysis of branching processes. Collaboration on projects concerning applications of branching processes to neutron transport.

Mobility program/type of mobility: Research stay.

Visited institution: University of Durham

Country: UK

Dates: 12/09/22 - 16/09/22

Context of the visit: Collaboration on genealogies of branching processes.

Mobility program/type of mobility: Research stay as part of PHC Allicance programme.

Visited institution: University of Auckland

Country: New Zealand

Dates: 07/11/22 - 19/11/22

Context of the visit: Collaboration on genealogies of branching processes. Workshop on branching processes

Mobility program/type of mobility: Research stay and workshop.

Visited institution: University of Melbourne

Country: Australia

Dates: 21/11/22 - 02/12/22

Context of the visit: Collaboration on parameter estimation for branching processes.

Mobility program/type of mobility: Research stay.

Visited institution: University of Durham

Country: UK

Dates: 19/12/22 - 22/12/22

Context of the visit: Collaboration on genealogies of branching processes.

Mobility program/type of mobility: Research stay as part of PHC Allicance programme.

8.3 European initiatives

8.3.1 Other european programs/initiatives

PHC Alliance 2022/23 E. Horton and Dr. Ellen Powell (University of Durham) obtained funding for bilateral mobility to work on applications of branching processes to the nuclear industry.

PHC FASIC Chercheur 2022 E. Horton and Dr. Sophie Hautphenne (University of Melbourne) obtained funding to work on parameter estimation of subcritical branching processes that arise in ecology.

Project PID2021-122442NB-I00 F. Dufour. "Analysis and control of deterministic, stochastic, and game theoretical dynamical systems" from 01/09/2022 to 31/08/2025. Ministry of Science and Innovation, Spain.

8.4 National initiatives

Naval Group Astral is a joint INRIA team project with Naval Group. The topic of this collaboration is described in section 4.1.

QuAMProcs of the program *Project Blanc* of the ANR The mathematical analysis of metastable processes started 75 years ago with the seminal works of Kramers on Fokker-Planck equation. Although the original motivation of Kramers was to « elucidate some points in the theory of the velocity of chemical reactions », it turns out that Kramers' law is observed to hold in many scientific fields: molecular biology (molecular dynamics), economics (modelization of financial bubbles), climate modeling, etc. Moreover, several widely used efficient numerical methods are justified by the mathematical description of this phenomenon.
Recently, the theory has witnessed some spectacular progress thanks to the insight of new tools coming from Spectral and Partial Differential Equations theory.
Semiclassical methods together with spectral analysis of Witten Laplacian gave very precise results on reversible processes. From a theoretical point of view, the semiclassical approach allowed to prove a complete asymptotic expansion of the small eigen values of Witten Laplacian in various situations (global problems, boundary problems, degenerate diffusions, etc.). The interest in the analysis of boundary problems was rejuvenated by recent works establishing links between the Dirichlet problem on a bounded domain and the analysis of exit event of the domain. These results open numerous perspectives of applications. Recent progress also occurred on the analysis of irreversible processes (e.g. on overdamped Langevin equation in irreversible context or full (inertial) Langevin equation).
The above progresses pave the way for several research tracks motivating our project: overdamped Langevin equations in degenerate situations, general boundary problems in reversible and irreversible case, non-local problems, etc.

Mission pour les initiatives transverses et interdisciplinaires, Défi Modélisation du Vivant, projet MISGIVING The aim of MISGIVING (MathematIcal Secrets penGuins dIVING) is to use mathematical models to understand the complexity of the multiscale decision process conditioning not only the optimal duration of a dive but also the diving behaviour of a penguin inside a bout. A bout is a sequence of succesive dives where the penguin is chasing prey. The interplay between the chasing period (dives) and the resting period due to the physiological cost of a dive (the time spent at the surface) requires some kind of optimization.

9 Dissemination

9.1 Promoting scientific activities

Participants: All team members.

9.1.1 Scientific events: organisation

Member of the organizing committees Pierrick Legrand was the co-organiser of the EA2022, Exeter, England.

9.1.2 Scientific events: selection

Chair of conference program committees Pierrick Legrand was chairman of the conference program committee of EA2022.

Reviewer All team members are regular reviewers for leading conferences in probability and applied statistics, control theory, signal processing and artificial evolution.

9.1.3 Journal

Member of the editorial boards P. Del Moral is an Associated Editor in the Annals of Applied Probability, since 2019.
P. Del Moral is an Associated Editor in Foundations of Data Science, since 2018.
P. Del Moral is an Associated Editor in Stochastic Analysis and Applications, since 2001.
F. Dufour was an associate editor for the SIAM Journal of Control and Optimization from 2009 to 2018.
F. Dufour is a corresponding editor for the SIAM Journal of Control and Optimization since 2018.
F. Dufour is an associate editor for the Journal Applied Mathematics and Optimization since 2018.
F. Dufour is an associate editor for the Journal Stochastics since 2018.
F. Dufour was the representative of the SIAM activity group in control and system theory for the journal SIAM News from 2014 to 2020.
F. Dufour was the chair of the selection committee for the 2021 SIAG/CST Best SICON Paper Prize.
F. Dufour is a member of the IFAC Technical Committee TC 1.4 Stochastic Systems.
J. Saracco is member of the Editorial Board of Astrostatistics (specialty section of Frontiers in Astronomy and Space Sciences).
M. Chavent is a member of the editorial board of the collection *Pratique R*, EDP Sciences.
Pierrick Legrand is the main editor for the Springer LNCS volumes *Artificial Evolution since 2009.*

Reviewer - reviewing activities *All team members are regular reviewers for leading journals in probability and applied statistics, control theory, signal processing and artificial evolution.*

9.1.4 Invited talks

Emma Horton

- *Probability seminar, KTH, Sweden, March 2022.*
- *Workshop: Pólya urns, stochastic approximation and quasi-stationary distributions: new developments, University of Bath, UK, April 2022.*
- *Workshop: Branching systems, reaction-diffusion equations and population models, Université de Montréal, Canada, May 2022.*
- *Probability seminar, University of Durham, UK, September 2022.*
- *Auckland Probability Workshop on Branching Processes, University of Auckland, New Zealand, November 2022.*
- *Asia-Pacific Seminar in Probability and Statistics, online seminar, November 2022.*

9.1.5 Leadership within the scientific community

Pierrick Legrand is the president of the association Evolution Artificielle.

9.1.6 Scientific expertise

E. Horton is secretary of the RSS Applied Probability section committee.
J. Saracco is regularly an expert for the HCERES (expertise of French mathematics laboratories) and for the ANRT (expertise of CIFRE thesis applications).
P. Legrand is regularly an expert for the ANR program.

9.1.7 Research administration

• National responsibilities
J. Saracco is treasurer of the French Statistical Society (SFdS) and an elected member of its board.
J. Saracco is an elected member of the national council of universities in applied mathematics (CNU 26).
• Universities and schools
F. Dufour is an invited member of the scientific council of the Institute of Mathematics of Bordeaux.
F. Dufour is member of the council of ENSEIRB MATMECA - Bordeaux INP since 2022.
J. Saracco is an elected member of the Bordeaux INP Board of Studies.
J. Saracco is an elected member of the Board of ENSC Bordeaux INP.
J. Saracco is head of the research team "OptimAl" (mathematical optimisation, random and statistical models) of the Mathematical Institute of Bordeaux (UMR 5251 CNRS).
M. Chavent is a member of the council of the department Sciences de l'ingénierie et du numérique of Bordeaux University.
M. Chavent is an elected member of the scientific council of the Institute of Mathematics of Bordeaux (since 2022).
A. Génadot has been an elected member of the scientific council of the Institute of Mathematics of Bordeaux (2019-2022).
P. Legrand has been an elected member of the laboratory council of the Institute of Mathematics of Bordeaux (2019-2022).
P. Legrand is an elected member of the scientific council of the Institute of Mathematics of Bordeaux (2022-).
P. Legrand is the president of the consultative commission in section CNU 26 of the Institute of Mathematics of Bordeaux (2019-).
• Inria
M. Chavent is a member of the CDT (Commission for Technological Development) at Inria Bordeaux since september 2022.
M. Chavent was a member of the Inria Evaluation Committee (2015-2019).
P. Del Moral is a member of the "Bureau du Comité des Projets" of the INRIA Bordeaux-Sud Ouest Research Center, since 2018.
A. Génadot is a member of the CER of Inria Bordeaux.
P. Legrand is a member of the CUMI commission.

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

- *J. Saracco is the head of the engineering department of ENSC, Graduate School of Cognitics (applied cognitive science and technology) which is a Bordeaux INP engineering school.*
- *Marie Chavent is in charge of the first year of the MIASHS degree program at Université de Bordeaux.*
- *Alexandre Genadot is in charge of the first year of the MIASHS degree program at Université de Bordeaux.*

- *Pierrick Legrand is in charge of the mathematics program for the MIASHS degree at Université de Bordeaux.*
- *Licence : P. Legrand, Algèbre, 129h, L1, Université de Bordeaux, France.*
- *Licence : P. Legrand, Espaces Euclidiens, 46,5h, L2, Université de Bordeaux, France.*
- *Licence : P. Legrand, Informatique pour les mathématiques, 30h, L2, Université de Bordeaux, France.*
- *DU : P. Legrand, Evolution Artificielle, Big data, 8h, DU, Bordeaux INP, France.*
- *Engineer School: Signal processing, 54 hours, ENSC, Bordeaux, France.*
- *Master: Scientific courses, 10 hours, Université de Bordeaux, France.*
- *Licence : A. Genadot, Bases en Probabilités, 18h, L1, Université de Bordeaux, France.*
- *Licence : A. Genadot, Projet Professionnel de l'étudiant, 8h, L1, Université de Bordeaux, France.*
- *Licence : A. Genadot, Probabilité, 30h, L2, Université de Bordeaux, France.*
- *Licence : A. Genadot, Techniques d'Enquêtes, 10h, L2, Université de Bordeaux, France.*
- *Licence : A. Genadot, Modélisation Statistiques, 16.5h, L3, Université de Bordeaux, France.*
- *Licence : A. Genadot, Préparation Stage, 15h, L3, Université de Bordeaux, France.*
- *Licence : A. Genadot, TER, 5h, L3, Université de Bordeaux, France.*
- *Licence : A. Genadot, Processus, 16.5h, L3, Université de Bordeaux, France.*
- *Licence : A. Genadot, Statistiques, 20h, L3, Bordeaux INP, France.*
- *Master : A. Genadot, Savoirs Mathématiques, 81h, M1, Université de Bordeaux et ESPE, France.*
- *Master : A. Genadot, Martingales, 29h, M1, Université de Bordeaux, France.*
- *Licence : F. Dufour, Probabilités et statistiques, 70h, first year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.*
- *Master : F. Dufour, Approche probabiliste et méthode de Monte Carlo, 24h, third year of école ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.*
- *Licence : J. Saracco, Probabilités et Statistique, 27h, first year of Graduate Schools of Engineering ENSC-Bordeaux INP, Institut Polytechnique de Bordeaux, France.*
- *Licence : J. Saracco, Statistique inférentielle et Analyse des données, 45h, first year of Graduate Schools of Engineering ENSC-Bordeaux INP, Institut Polytechnique de Bordeaux, France.*
- *Licence : J. Saracco, Statistique pour l'ingénieur, 16h, first year of Graduate Schools of Engineering ENSPIMA-Bordeaux INP, Institut Polytechnique de Bordeaux, France.*
- *Master : J. Saracco, Modélisation statistique, 81h, second year of Graduate Schools of Engineering ENSC-Bordeaux INP, Institut Polytechnique de Bordeaux, France.*
- *DU : J. Saracco, Statistique et Big data, 45h, DU BDSI (Big data et statistique pour l'ingénieur), Bordeaux INP, France.*
- *Licence : M. Chavent, Statistique Inférentielle, 18h, L2, Université de Bordeaux, France*
- *Licence : M. Chavent, Techniques d'Enquêtes, 10h, L2, Université de Bordeaux, France*
- *Master : M. Chavent, DataMining, 43h, M2, Université de Bordeaux*
- *Master : M. Chavent, Machine Learning, 58h, Université de Bordeaux,*

- *DU: M. Chavent, Apprentissage, 12h, DU BDSI, Bordeaux INP, France*
- *Licence : E. Horton, Probabilités et Statistiques, 32h, ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.*
- *Licence : E. Horton, Probabilités, 22h, ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.*
- *Licence : E. Horton, Probabilités, 24h, ENSEIRB-MATMECA, Institut Polytechnique de Bordeaux, France.*
- *Licence : E. Horton, 30h, Statistique inférentielle et Analyse des données, first year of Graduate Schools of Engineering ENSC-Bordeaux INP, Institut Polytechnique de Bordeaux, France.*

9.2.2 Supervision

- *E. Horton co-supervised a summer project with Prof. Alexander Cox (University of Bath) on sensitivity analysis of branching processes.*

9.2.3 Juries

Pierrick Legrand was member of the jury of the Phd of Yann Cabanes.

10 Scientific production

10.1 Major publications

[1] A. M. G. Cox, E. L. Horton, A. E. Kyprianou and D. Villemonais. 'Stochastic Methods for Neutron Transport Equation III: Generational many-to-one and $k_{\mathtt{eff}}$'. In: *SIAM Journal on Applied Mathematics* 81.3 (27th May 2021). DOI: 10.1137/19M1295854. URL: https://hal.archives-ouvertes.fr/hal-02390745.

[2] P. Del Moral and E. Horton. *A theoretical analysis of one-dimensional discrete generation ensemble Kalman particle filters.* 3rd July 2021. URL: https://hal.inria.fr/hal-03277374.

[3] F. Dufour and T. Prieto-Rumeau. *Stationary Markov Nash equilibria for nonzero-sum constrained ARAT Markov games.* 4th Jan. 2022. URL: https://hal.inria.fr/hal-03510818.

[4] M.-P. Ellies-Oury, D. Durand, A. Listrat, M. Chavent, J. Saracco and D. Gruffat. 'Certain relationships between Animal Performance, Sensory Quality and Nutritional Quality can be generalized between various experiments on animal of similar types'. In: *Livestock Science* 250 (Aug. 2021), p. 104554. DOI: 10.1016/j.livsci.2021.104554. URL: https://hal.archives-ouvertes.fr/hal-03272625.

[5] A. Genadot. 'Quina metodologia per despolhar l'enquèsta Bourciez ?' In: Ièr, deman : diga-m'o dins la lenga. Montpellier, France, 26th Nov. 2021. URL: https://hal.inria.fr/hal-03471470.

[6] S. Girard, H. Lorenzo and J. Saracco. 'Advanced topics in Sliced Inverse Regression'. In: *Journal of Multivariate Analysis* 188 (2022), p. 104852. DOI: 10.1016/j.jmva.2021.104852. URL: https://hal.inria.fr/hal-03367798.

[7] E. Grivel, B. Berthelot, P. Legrand and A. Giremus. 'DFA-based abacuses providing the Hurst exponent estimate for short-memory processes'. In: *Digital Signal Processing* 116 (2021). DOI: 10.1016/j.dsp.2021.103102. URL: https://hal.archives-ouvertes.fr/hal-03225784.

[8] H. Lorenzo, O. Cloarec, R. Thiébaut and J. Saracco. 'Data-Driven Sparse Partial Least Squares'. In: *Statistical Analysis and Data Mining* (25th Dec. 2021). URL: https://hal.inria.fr/hal-03368956.

[9] A. Saadoun, A. Schein, V. Péan, P. Legrand, L. S. Aho Glélé and A. Bozorg Grayeli. 'Frequency Fitting Optimization Using Evolutionary Algorithm in Cochlear Implant Users with Bimodal Binaural Hearing'. In: *Brain Sciences* 12.2 (Feb. 2022), p. 253. DOI: 10.3390/brainsci12020253. URL: https://hal.science/hal-03610651.

[10] T. Vanhatalo, P. Legrand, M. Desainte-Catherine, P. Hanna, A. Brusco, G. Pille and Y. Bayle. 'A Review of Neural Network-Based Emulation of Guitar Amplifiers'. In: *Applied Sciences* 12.12 (June 2022), p. 5894. DOI: 10.3390/app12125894. URL: https://hal.inria.fr/hal-03881859.

10.2 Publications of the year

International journals

[11] M. Chavent, V. Kuentz-Simonet, A. Labenne and J. Saracco. 'Multivariate Analysis of Mixed Data: The R Package PCAmixdata'. In: *Electronic Journal of Applied Statistical Analysis* (2022). URL: https://hal.archives-ouvertes.fr/hal-01662595.

[12] D. Crisan, P. del Moral, A. Jasra and H. Ruzayqat. 'Log-normalization constant estimation using the ensemble Kalman–Bucy filter with application to high-dimensional models'. In: *Advances in Applied Probability* 54.4 (Dec. 2022), pp. 1139–1163. DOI: 10.1017/apr.2021.62. URL: https://hal.inria.fr/hal-03850299.

[13] F. Dufour, A. Piunovskiy and A. Plakhov. 'On the equivalence of the integral and differential Bellman equations in impulse control problems'. In: *International Journal of Control, Automation and Systems* (2022). URL: https://hal.inria.fr/hal-03507200.

[14] F. Dufour and T. Prieto-Rumeau. 'Maximizing the probability of visiting a set infinitely often for a countable state space Markov decision process'. In: *Journal of Mathematical Analysis and Applications* (2022). DOI: 10.1016/j.jmaa.2021.125639. URL: https://hal.inria.fr/hal-03507261.

[15] F. Dufour and T. Prieto-Rumeau. 'Stationary Markov Nash equilibria for nonzero-sum constrained ARAT Markov games'. In: *SIAM Journal on Control and Optimization* (2022). DOI: 10.1137/21M144565X. URL: https://hal.inria.fr/hal-03510818.

[16] S. Girard, H. Lorenzo and J. Saracco. 'Advanced topics in Sliced Inverse Regression'. In: *Journal of Multivariate Analysis* 188 (2022), p. 104852. DOI: 10.1016/j.jmva.2021.104852. URL: https://hal.inria.fr/hal-03367798.

[17] I. Gonzalez, E. Horton and A. E. Kyprianou. 'Asymptotic moments of spatial branching processes'. In: *Probability Theory and Related Fields* 184 (25th Apr. 2022), pp. 805–858. DOI: 10.1007/s00440-022-01131-2. URL: https://hal.archives-ouvertes.fr/hal-03321803.

[18] E. Grivel, J. Soetens, G. Colin and P. Legrand. 'Combining the global trends of DFA or CDFA of different orders'. In: *Digital Signal Processing* (2023). URL: https://hal.archives-ouvertes.fr/hal-03920420.

[19] S. C. Harris, E. Horton, A. E. Kyprianou and M. Wang. 'Yaglom limit for critical nonlocal branching markov processes'. In: *Annals of Probability* 50.6 (Nov. 2022), pp. 2373–2408. DOI: 10.1214/22-AOP1585. URL: https://hal.inria.fr/hal-03276237.

[20] E. Horton and A. R. Watson. 'Strong laws of large numbers for a growth-fragmentation process with bounded cell sizes'. In: *ALEA : Latin American Journal of Probability and Mathematical Statistics* (2022). URL: https://hal.inria.fr/hal-03276236.

[21] P. del Moral and E. Horton. 'A note on Riccati matrix difference equations'. In: *SIAM Journal on Control and Optimization* (2022). URL: https://hal.inria.fr/hal-03299378.

[22] P. del Moral and E. Horton. 'A theoretical analysis of one-dimensional discrete generation ensemble Kalman particle filters'. In: *Annals of Applied Probability* (2022). URL: https://hal.inria.fr/hal-03277374.

[23] P. del Moral, E. Horton and A. Jasra. 'On the Stability of Positive Semigroups'. In: *Annals of Applied Probability* (2022). URL: https://hal.inria.fr/hal-03468416.

[24] T. Vanhatalo, P. Legrand, M. Desainte-Catherine, P. Hanna, A. Brusco, G. Pille and Y. Bayle. 'A Review of Neural Network-Based Emulation of Guitar Amplifiers'. In: *Applied Sciences* 12.12 (June 2022), p. 5894. DOI: 10.3390/app12125894. URL: https://hal.inria.fr/hal-03881859.

International peer-reviewed conferences

[25] T. Vanhatalo, P. Legrand, M. Desainte-Catherine, P. Hanna, A. Brusco, G. Pille and Y. Bayle. 'Neural Network-based Virtual Analog Modelling'. In: Artificial Evolution 2022. Exeter, United Kingdom, 31st Oct. 2022. URL: https://hal.inria.fr/hal-03852692.

Conferences without proceedings

[26] M. Chavent, J. Lacaille, A. Mourer and M. Olteanu. 'Sparse and group-sparse clustering for mixed data An illustration of the vimpclust package'. In: JDS 2022 - 53èmes Journées de Statistique de la Société Française de Statistique (SFdS). Lyon, France, 13th June 2022. URL: https://hal.archives-ouvertes.fr/hal-03839521.

[27] H. Lorenzo, J. Saracco and O. Cloarec. 'Imputation for supervised learning problems in high dimension'. In: 24th International Conference on Computational Statistics (COMPSTAT 2022), 23-26 August 2022, Bologna (Italy). Bologna, Italy, 2022. URL: https://hal.archives-ouvertes.fr/hal-03842462.

[28] J. Saracco and H. Lorenzo. 'Détection d'individus atypiques en régression SIR'. In: Dataquitaine 2022 (IA, Recherche Opérationnelle & Data Science), Février 2022, Bordeaux, France. Bordeaux, France, Feb. 2022. URL: https://hal.archives-ouvertes.fr/hal-03842500.

[29] J. Saracco and H. Lorenzo. 'Sélection de variables en régression SIR par seuillage doux ou seuillage dur de la matrice d'intérêt'. In: Journées de Statistique 2022 – Société Française de Statistique, Juin 2022, Lyon, France. Lyon, France, June 2022. URL: https://hal.archives-ouvertes.fr/hal-03842492.

[30] J. Saracco, H. Lorenzo and O. Cloarec. 'Koh-Lanta, missing data imputation in supervised context'. In: 23ème Congrès de Chimiométrie. Brest, France, June 2022. URL: https://hal.archives-ouvertes.fr/hal-03842483.

[31] J. Saracco, H. Lorenzo and S. Girard. 'Advanced topics in Sliced Inverse Regression V2'. In: JMVA 2022 - 50th Anniversary Jubilee. Virtual Meeting, United States, 20th Apr. 2022. URL: https://hal.archives-ouvertes.fr/hal-03842478.

Scientific books

[32] M. Chavent, M. Olteanu, M. Cottrell, J. Lacaille and A. Mourer. *Sparse Weighted K-Means for Groups of Mixed-Type Variables*. Vol. 533. Lecture Notes in Networks and Systems. Springer International Publishing, 2022. DOI: 10.1007/978-3-031-15444-7. URL: https://hal.archives-ouvertes.fr/hal-03827823.

Scientific book chapters

[33] O. Costa and F. Dufour. 'Optimal Control of Piecewise Deterministic Markov Processes'. In: *Stochastic Analysis, Filtering, and Stochastic Optimization*. Springer International Publishing, 22nd Apr. 2022, pp. 53–77. DOI: 10.1007/978-3-030-98519-6_3. URL: https://hal.inria.fr/hal-03789468.

Doctoral dissertations and habilitation theses

[34] A. Mourer. 'Variable selection and importance in statistical learning: Applications to highly correlated aircraft engine test data'. Paris 1 - Panthéon-Sorbonne, 12th Apr. 2022. URL: https://hal.archives-ouvertes.fr/tel-03842117.

Reports & preprints

[35] A. Aleksian, P. del Moral, A. Kurtzmann and J. Tugaut. *On the exit-problem for self-interacting diffusions V2*. 13th Nov. 2022. URL: https://hal.inria.fr/hal-03850314.

[36] A. M. Cox, E. Horton and D. Villemonais. *Binary branching processes with Moran type interactions*. 7th July 2022. URL: https://hal.inria.fr/hal-03715678.

[37] F. Dufour and O. L. D. V. Costa. *Adaptive Discounted Control for Piecewise Deterministic Markov Processes*. 15th Nov. 2022. URL: https://hal.archives-ouvertes.fr/hal-03853095.

[38] F. Dufour and T. Prieto-Rumeau. *Maximizing the probability of visiting a set infinitely often for a Markov decision process*. 15th Nov. 2022. URL: https://hal.archives-ouvertes.fr/hal-03853326.

[39] A. Genadot. *Averaging for slow-fast piecewise deterministic Markov processes with an attractive boundary*. 18th Jan. 2022. URL: https://hal.inria.fr/hal-03531257.

[40] S. C. Harris, E. Horton, A. E. Kyprianou and E. Powell. *Many-to-few for non-local branching Markov process*. 16th Nov. 2022. URL: https://hal.inria.fr/hal-03854678.

10.3 Cited publications

[41] E. Altman. *Constrained Markov decision processes*. Stochastic Modeling. Chapman & Hall/CRC, Boca Raton, FL, 1999, pp. xii+242.

[42] N. Bäuerle and U. Rieder. *Markov decision processes with applications to finance*. Universitext. Springer, Heidelberg, 2011, pp. xvi+388. DOI: 10.1007/978-3-642-18324-9. URL: https://doi.org/10.1007/978-3-642-18324-9.

[43] D. P. Bertsekas and S. E. Shreve. *Stochastic optimal control: The discrete time case*. Vol. 139. Mathematics in Science and Engineering. New York: Academic Press Inc., 1978, pp. xiii+323.

[44] O. L. d. V. Costa and F. Dufour. *Continuous average control of piecewise deterministic Markov processes*. SpringerBriefs in Mathematics. Springer, New York, 2013, pp. xii+116. DOI: 10.1007/978-1-4614-6983-4. URL: https://doi.org/10.1007/978-1-4614-6983-4.

[45] M. H. A. Davis. *Markov models and optimization*. Vol. 49. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1993, pp. xiv+295. DOI: 10.1007/978-1-4899-4483-2. URL: http://dx.doi.org/10.1007/978-1-4899-4483-2.

[46] P. Del Moral. *Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer-Verlag, New York, 2004, p. 573.

[47] P. Del Moral. *Mean field simulation for Monte Carlo integration*. Monographs on Statistics and Applied Probability. Chapman and Hall, 2013. URL: http://www.crcpress.com/product/isbn/9781466504059.

[48] P. Del Moral, A. Doucet and J. A. 'Sequential Monte Carlo samplers'. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68.3 (2006), pp. 411–436.

[49] P. Del Moral and A. Guionnet. 'On the stability of interacting processes with applications to filtering and genetic algorithms'. In: Annales de l'Institut Henri Poincaré 37.2 (2001), pp. 155–194.

[50] P. Del Moral and A. Guionnet. 'On the stability of Measure Valued Processes with Applications to filtering'. In: C.R. Acad. Sci. Paris, t. 329, Serie I (1999), pp. 429–434.

[51] P. Del Moral, J. Jacod and P. Protter. 'The Monte-Carlo method for filtering with discrete-time observations'. In: Probability Theory and Related Fields 120.3 (2001), pp. 346–368.

[52] P. Del Moral and L. Miclo. *Branching and Interacting Particle Systems Approximations of Feynman-Kac Formulae with Applications to Non-Linear Filtering*. Vol. 1729. Séminaire de Probabilités XXXIV. Ed. J. Azéma et al., 2000, pp. 1–145.

[53] P. Del Moral and S. Penev. *Stochastic Processes: From Applications to Theory*. Chapman and Hall/CRC, 2017.

[54] P. Del Moral and J. Tugaut. 'On the stability and the uniform propagation of chaos properties of ensemble Kalman-Bucy filters'. In: Annals of Applied Probability 28.2 (2018), pp. 790–850.

[55] P. Del Moral. 'Non Linear Filtering: Interacting Particle Solution'. In: Markov Processes and Related Fields 2.4 (1996), pp. 555–580.

[56] E. Dynkin and A. Yushkevich. *Controlled Markov processes.* Vol. 235. Grundlehren der Mathematischen Wissenschaften. Berlin: Springer-Verlag, 1979, pp. xvii+289.

[57] J. Filar and K. Vrieze. *Competitive Markov decision processes.* New York: Springer-Verlag, 1997, pp. xii+393.

[58] X. Guo and O. Hernández-Lerma. *Continuous-time Markov decision processes.* Vol. 62. Stochastic Modelling and Applied Probability. Theory and applications. Springer-Verlag, Berlin, 2009, pp. xviii+231. DOI: 10.1007/978-3-642-02547-1. URL: https://doi.org/10.1007/978-3-642-02547-1.

[59] O. Hernández-Lerma. *Adaptive Markov control processes.* Vol. 79. Applied Mathematical Sciences. New York: Springer-Verlag, 1989, pp. xiv+148.

[60] O. Hernández-Lerma and J.-B. Lasserre. *Discrete-time Markov control processes: Basic optimality criteria.* Vol. 30. Applications of Mathematics. New York: Springer-Verlag, 1996, pp. xiv+216.

[61] O. Hernández-Lerma and J.-B. Lasserre. *Further topics on discrete-time Markov control processes.* Vol. 42. Applications of Mathematics. New York: Springer-Verlag, 1999, pp. xiv+276.

[62] K. Hinderer. *Foundations of non-stationary dynamic programming with discrete time parameter.* Lecture Notes in Operations Research and Mathematical Systems, Vol. 33. Springer-Verlag, Berlin-New York, 1970, pp. vi+160.

[63] A. Hordijk and F. van der Duyn Schouten. 'Markov decision drift processes: conditions for optimality obtained by discretization'. In: *Math. Oper. Res.* 10.1 (1985), pp. 160–173. DOI: 10.1287/moor.10.1.160. URL: https://doi.org/10.1287/moor.10.1.160.

[64] A. Hordijk and F. A. van der Duyn Schouten. 'Discretization and weak convergence in Markov decision drift processes'. In: *Math. Oper. Res.* 9.1 (1984), pp. 112–141. DOI: 10.1287/moor.9.1.112. URL: http://dx.doi.org/10.1287/moor.9.1.112.

[65] V. kolokoltsov. 'Nonlinear Markov Processes and Kinetic Equations'. In: Cambridge Univ. Press (2010).

[66] A. B. Piunovskiy. *Examples in Markov decision processes.* Vol. 2. Imperial College Press Optimization Series. Imperial College Press, London, 2013, pp. xiv+293.

[67] A. B. Piunovskiy. *Optimal control of random sequences in problems with constraints.* Vol. 410. Mathematics and its Applications. With a preface by V. B. Kolmanovskii and A. N. Shiryaev. Kluwer Academic Publishers, Dordrecht, 1997, pp. xii+345. DOI: 10.1007/978-94-011-5508-3. URL: https://doi.org/10.1007/978-94-011-5508-3.

[68] T. Prieto-Rumeau and O. Hernández-Lerma. *Selected topics on continuous-time controlled Markov chains and Markov games.* Vol. 5. ICP Advanced Texts in Mathematics. Imperial College Press, London, 2012, pp. xii+279. DOI: 10.1142/p829. URL: https://doi.org/10.1142/p829.

[69] M. Puterman. *Markov decision processes: discrete stochastic dynamic programming.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc., 1994, pp. xx+649.