RESEARCH CENTRE

Inria Centre at the University of Bordeaux

IN PARTNERSHIP WITH:

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

2024

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

Contents

Project-Team ASTRAL **1**

1 Team members, visitors, external collaborators **3**

2 Overall objectives **4**
2.1 Outline of the research project 4
2.2 Approach and methodologies 5
2.3 Innovation and industrial transfer 6

3 Research program **7**
3.1 Statistical learning 7
3.2 Stochastic learning 9
3.3 Decision and stochastic control 11

4 Application domains **13**
4.1 Naval Group research activities 13
4.2 Other collaborations 14

5 Highlights of the year **15**

6 New software, platforms, open data **15**
6.1 New software 15
6.1.1 FracLab 15
6.1.2 PCAmixdata 16
6.1.3 vimplclust 16
6.1.4 divdiss 16

7 New results **16**
7.1 Statistical learning 16
7.1.1 Modelling the physiological, muscular, and sensory characteristics in relation to beef quality from 15 cattle breeds 16
7.1.2 A two-sample tree-based test for hierarchically organized genomic signals 17
7.2 Stochastic learning 17
7.2.1 Bayesian Parameter Inference for Partially Observed Diffusions using Multilevel Stochastic Runge-Kutta Methods 17
7.2.2 Geometric Convergence and Concentration Inequalities for the Feynman–Kac Genetic Algorithm 18
7.2.3 Self-interacting diffusions: long-time behaviour and exit-problem in the convex case 18
7.3 Stochastic control and games 18
7.3.1 Adaptive average control for piecewise deterministic Markov processes 18
7.3.2 Minimum Contrast Estimators for Piecewise Deterministic Markov Processes 18
7.3.3 Absorbing Markov decision processes and their occupation measures. 19
7.3.4 Nash equilibria for total expected reward absorbing Markov games: the constrained and unconstrained cases. 19
7.3.5 Optimal stochastic control and application on passive trajectography 19
7.4 Signal processing, artifical evolution and neural networks 20
7.4.1 Benefits of Zero-Phase or Linear Phase Filters to Design Multiscale Entropy: Theory and Application 20
7.4.2 Reservoir Computing for Short High-Dimensional Time Series: an Application to SARS-CoV-2 Hospitalization Forecast 20
7.4.3 Real-time simulation of nonlinear audio effects using artificial intelligence 21
7.4.4 Tackling Long-Range Dependencies in Dynamic Range Compression Modeling via Deep Learning 21

8 **Bilateral contracts and grants with industry** 22
8.1 Bilateral contracts with industry . . . 22
8.2 Bilateral Grants with Industry . . . 22

9 **Partnerships and cooperations** 23
9.1 International initiatives . . . 23
9.1.1 Participation in other International Programs . . . 23
9.2 International research visitors . . . 23
9.2.1 Visits of international scientists . . . 23
9.3 European initiatives . . . 24
9.3.1 Other european programs/initiatives . . . 24
9.4 National initiatives . . . 24

10 **Dissemination** 24
10.1 Promoting scientific activities . . . 25
10.1.1 Scientific events: organisation . . . 25
10.1.2 Scientific events: selection . . . 25
10.1.3 Journal . . . 25
10.1.4 Invited talks . . . 25
10.1.5 Scientific expertise . . . 26
10.2 Teaching - Supervision - Juries . . . 26
10.2.1 Teaching . . . 26
10.2.2 Supervision . . . 27
10.2.3 Juries . . . 27
10.3 Popularization . . . 27
10.3.1 Participation in Live events . . . 27

11 **Scientific production** 27
11.1 Major publications . . . 27
11.2 Publications of the year . . . 28
11.3 Cited publications . . . 30

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 Scientist

- Pierre Del Moral [INRIA, Senior Researcher]

Faculty Members

- François Dufour [Team leader, BORDEAUX INP, Professor Delegation, HDR]
- Marie Chavent [UNIV BORDEAUX, Professor, HDR]
- Alexandre Genadot [UNIV BORDEAUX, Associate Professor, HDR]
- Pierrick Legrand [UNIV BORDEAUX, Associate Professor, until Aug 2024, HDR]
- Pierrick Legrand [BORDEAUX INP, Professor, from Sep 2024, HDR]
- Jérôme Saracco [BORDEAUX INP, Professor, HDR]

PhD Students

- Yann Bourdin [ARTURIA, CIFRE]
- Luc De Montella [NAVAL GROUP, CIFRE]
- Mariette Dupuy [UNIV BORDEAUX]
- Romain Namyst [INRIA]
- Valentin Portmann [INRIA]
- Tara Vanhatalo [OROSYS, until Apr 2024]

Technical Staff

- Denis Arrivault [INRIA, Engineer]
- Enzo Iglesis [INRIA, Engineer]
- Dann Laneuville [NAVAL GROUP, Engineer]
- Olivier Marceau [NAVAL GROUP, Engineer]
- Adrien Negre [NAVAL GROUP, Engineer]

Interns and Apprentices

- Noah Kollin [INRIA, Intern, from Mar 2024 until Jun 2024]

Administrative Assistant

- Ellie Correa Da Costa De Castro Pinto [INRIA]

External Collaborators

- Alexander Cox [Department of Mathematical Sciences, University of Bath]
- Oswaldo Do Valle Costa [Escola Politécnica da USP Departamento de Engenharia de Telecomunicações e Controle]
- Oscar Key [UCL in the Centre for Artificial Intelligence]
- Bastien Mallein [Université Toulouse III Paul Sabatier, IMT.]
- Pierre Kristoffer Nyquist [KTH Royal Institute of Technology]
- Alexander Watson [Department of Statistical Science at University College London]

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, 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, D. Laneuville, P. Legrand, A. Nègre, J. Saracco.

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 [39], [40], [46],[58] 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 [48], 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 [43, 42, 45] and in the more recent article [47]. The first use of particle algorithms and Approximated Bayesian Computation type methodologies in nonlinear filtering seems to have started in [44]. Last but not least, the development of sequential Monte Carlo methodology in statistics was introduced in the seminal article [41].

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 [48, 44, 43], see also the books [45, 39, 40, 46] 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 [48] 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, D. Laneuville, O. Marceau, A. Nègre, J. Saracco.

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 [34, 36, 49, 50, 52, 53, 54, 55, 60, 59, 62] 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) [35, 37, 38, 51, 57, 56, 61]. 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. 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, D. Laneuville, O. Marceau, A. Nègre.
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

Organization of the EA2024 conference in the Inria center at the University of Bordeaux at the end of October 2024.

Pierrick Legrand obtained a position as University Professor in 2024.

Tara Vanhatalo's PhD thesis defense

Romain Namyst's PhD thesis defense

6 New software, platforms, open data

6.1 New software

6.1.1 FracLab

Keyword: Stochastic process

Functional Description: FracLab is a general purpose signal and image processing toolbox based on fractal, multifractal and local regularity methods. FracLab can be approached from two different perspectives: - (multi-) fractal and local regularity analysis: A large number of procedures allow to compute various quantities associated with 1D or 2D signals, such as dimensions, Hölder and 2-microlocal exponents or multifractal spectra.

- Signal/Image processing: Alternatively, one can use FracLab directly to perform many basic tasks in signal processing, including estimation, detection, denoising, modeling, segmentation, classification, and synthesis.

URL: https://project.inria.fr/fraclab/

Contact: Jacques Levy-Vehel

Participants: Antoine Echelard, Christian Choque Cortez, Jacques Levy-Vehel, Khalid Daoudi, Olivier Barriere, Paulo Goncalves, Pierrick Legrand

Partners: Centrale Paris, Mas

6.1.2 PCAmixdata

Keyword: Statistic analysis

Functional Description: 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). The MFAmix procedure handles a mixture of numerical and categorical variables within a group - something which was not possible in the standard MFA procedure. We also included techniques to project new observations onto the principal components of the three methods in the new version of the package.

URL: https://cran.r-project.org/web/packages/PCAmixdata/index.html

Contact: Marie Chavent

6.1.3 vimplclust

Keywords: Clustering, Fair and ethical machine learning

Functional Description: vimpclust is an R package that implements methods related to sparse clustering and variable importance. The package currently allows to perform sparse k-means clustering with a group penalty, so that it automatically selects groups of numerical features. It also allows to perform sparse clustering and variable selection on mixed data (categorical and numerical features), by preprocessing each categorical feature as a group of numerical features. Several methods for visualizing and exploring the results are also provided.

URL: https://CRAN.R-project.org/package=vimpclust

Contact: Marie Chavent

6.1.4 divdiss

Name: divisive monothetic clustering on dissimilarity matrix

Keywords: Clustering, Machine learning

Functional Description: The div_diss function implements a divisive monotopic hierarchical classification algorithm.

URL: https://github.com/chavent/divdiss

Contact: Marie Chavent

7 New results

7.1 Statistical learning

7.1.1 Modelling the physiological, muscular, and sensory characteristics in relation to beef quality from 15 cattle breeds

Sensory beef quality is an important parameter for consumers, but it is difficult to understand what determines quality because of the large variation due to breed, type of animal, sex, and farm management. In [12], to address these challenges, meat samples from Longissimus thoracis (LT) were collected 24 h after slaughter from a total of 436 young bulls representing 15 cattle breeds. The samples were analysed for physiological, muscular, and sensory characteristics to evaluate beef quality. Ten groups of variables were identified by hierarchical cluster analysis of individual variables and Principal Component Analysis (PCA). The variables described distinct characteristics, including Physiological traits (animal maturity, growth rate, muscle mass), Sensory traits (tenderness, juiciness, flavor, meat colour), and some Muscular characteristics such as lipid content, maturation enzymes and oxidative metabolism, which are likely to discriminate among breeds. The analyses showed that muscular characteristics are positively associated with beef sensory characteristics whereas physiological characteristics are negatively associated with sensory quality. Dairy breeds produced beef rich in lipids with a strong flavour, French breeds showed fast growth rate and highly flavoured meat, and Italian breeds were characterised by good muscular development. In contrast, British breeds have oxidative muscles and produce beef with a strong flavour. Danish breeds have an intermediate score for the sensorial and muscular characteristics.

Participants: J. Albechaalany, M-P. Ellies-Oury, J. Saracco (ASTRAL), M.M. Campo, I. Richardson, P. Ertbjerg, S. Failla, B. Panea, J.L. Williams, M. Christensen, J.-F. Hocquette.

7.1.2 A two-sample tree-based test for hierarchically organized genomic signals

The article [22] addresses a common type of data encountered in genomic studies, where a signal along a linear chromosome exhibits a hierarchical organization. We propose a novel framework to assess the significance of dissimilarities between two sets of genomic matrices obtained from distinct biological conditions. Our approach relies on a data representation based on trees. It utilizes tree distances and an aggregation procedure for tests performed at the level of leaf pairs. Numerical experiments demonstrate its statistical validity and its superior accuracy and power compared to alternatives. The method's effectiveness is illustrated using real-world data from GWAS and Hi-C data.

Participants: Pierre Neuvial, Nathanaël Randriamihamison, Marie Chavent (ASTRAL), Sylvain Foissac, Nathalie Vialaneix.

7.2 Stochastic learning

7.2.1 Bayesian Parameter Inference for Partially Observed Diffusions using Multilevel Stochastic Runge-Kutta Methods

In [20], we consider the problem of Bayesian estimation of static parameters associated to a partially and discretely observed diffusion process. We assume that the exact transition dynamics of the diffusion process are unavailable, even up-to an unbiased estimator and that one must time-discretize the diffusion process. In such scenarios it has been shown how one can introduce the multilevel Monte Carlo method to reduce the cost to compute posterior expected values of the parameters for a pre-specified mean square error (MSE). These afore-mentioned methods rely on upon the Euler-Maruyama discretization scheme which is well-known in numerical analysis to have slow convergence properties. We adapt stochastic Runge-Kutta (SRK) methods for Bayesian parameter estimation of static parameters for diffusions. This can be implemented in high-dimensions of the diffusion and seemingly under-appreciated in the uncertainty quantification and statistics fields. For a class of diffusions and SRK methods, we consider the estimation of the posterior expectation of the parameters. We prove that to achieve a MSE of $\theta(\epsilon^2)$, for $\epsilon > 0$ given, the associated work is $\theta(\epsilon^{-2})$. Whilst the latter is achievable for the Milstein scheme, this method is often not applicable for diffusions in dimension larger than two. We also illustrate our methodology in several numerical examples.

Participants: Pierre del Moral (ASTRAL), Shulan Hu, Ajay Jasra, Hamza Ruzayqat, Xinyu Wang.

7.2.2 Geometric Convergence and Concentration Inequalities for the Feynman–Kac Genetic Algorithm

In [21], we consider a genetic evolution model associated to a given Feynman–Kac flow (called also the simple genetic algorithm). We first obtain an estimate of the contraction coefficient of this interacting particle system in some suitable metric, independent of the number of particles in the system. Second, by transport-entropy inequality technique, we obtain some concentration inequalities for the particle system, uniform in time and in the number of particles.

Participants: Pierre del Moral (ASTRAL), Xinyu Wang.

7.2.3 Self-interacting diffusions: long-time behaviour and exit-problem in the convex case

In [13], we study a class of time-inhomogeneous diffusion: the self-interacting one. We show a convergence result with a rate of convergence that does not depend on the diffusion coefficient. Finally, we establish a so-called Kramers' type law for the first exit-time of the process from domain of attractions when the landscapes are uniformly convex.

Participants: Ashot Aleksian, Pierre del Moral (ASTRAL), Aline Kurtzmann, Julian Tugaut.

7.3 Stochastic control and games

7.3.1 Adaptive average control for piecewise deterministic Markov processes

The main goal of the work presented in [14] is to study the adaptive infinite-horizon average continuous-time optimal control problem of piecewise deterministic Markov processes (PDMPs) for the case in which there are no boundary jumps. It is assumed that the control acts continuously on the jump intensity λ and on the transition measure Q of the process, and that jump parameters (λ and Q), as well as the cost C , depend on an unknown parameter β^* . It is shown that the principle of estimation and control holds, that is, the strategy consisting of choosing, at each stage n , an action according to an optimal stationary policy, where the true but unknown parameter β^* is replaced by its estimated value β^*_n , is average optimal, provided that the sequence of estimators $\{\beta^*_n\}$ of β^* is strongly consistent, that is, $\{\beta^*_n\}$ converge to β^* almost surely.

Participants: O.L.V Costa, François Dufour (ASTRAL), Alexandre Genadot (ASTRAL).

7.3.2 Minimum Contrast Estimators for Piecewise Deterministic Markov Processes

The main goal of the paper [15] is to study the minimum contrast estimator (MCE) approach for the parameter estimation problem of piecewise deterministic Markov processes (PDMPs), associated to adaptive control problems. It is assumed that the control acts continuously on the jump intensity λ and on the transition measure Q of the process, as well as on the costs, and that these parameters depend on an unknown parameter β^* . One of our objective is to introduce a minimum contrast estimator $(\beta^*_n)_{n\in N}$ for the family of PDMPs. Sufficient conditions are then presented to ensure that $(\beta^*_n)_{n\in N}$ is a strongly consistent estimator of β^* . It should be noticed that PDMPs are characterized by a deterministic motion punctuated by random jumps (either spontaneous or due to the flow touching a boundary), which brings new challenges in the analysis of the problem. The paper is concluded with a numerical example for the adaptive discounted control of PDMPs.

Participants: O.L.V Costa, François Dufour (ASTRAL), Alexandre Genadot (ASTRAL).

7.3.3 Absorbing Markov decision processes and their occupation measures.

In [16], we consider an absorbing Markov decision process with Borel state and action spaces. We study conditions under which the MDP is uniformly absorbing and the set of occupation measures of the MDP is compact in the usual weak topology. These include suitable continuity requirements on the transition kernel and conditions on the dynamics of the system at the boundary of the absorbing set. We generalize previously known results and give an answer to some conjectures that have been mentioned in the related literature.

In [17], we study discrete-time absorbing Markov Decision Processes (MDP) with measurable state space and Borel action space with a given initial distribution. For such models, solutions to the characteristic equation that are not occupation measures may exist. Several necessary and sufficient conditions are provided to guarantee that any solution to the characteristic equation is an occupation measure. Under the so-called continuity-compactness conditions, it is shown that the set of occupation measures is compact in the weak-strong topology if and only if the model is uniformly absorbing. Finally, it is shown that the occupation measures are characterized by the characteristic equation and an additional condition. Several examples are provided to illustrate our results.

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

7.3.4 Nash equilibria for total expected reward absorbing Markov games: the constrained and unconstrained cases.

In [18], we consider a nonzero-sum N-player Markov game on an abstract measurable state space with compact metric action spaces. The payoff functions are bounded Carathéodory functions and the transitions of the system are assumed to have a density function satisfying some continuity conditions. The optimality criterion of the players is given by a total expected payoff on an infinite discrete-time horizon. Under the condition that the game model is absorbing, we establish the existence of Markov strategies that are a noncooperative equilibrium in the family of all history-dependent strategies of the players for both the constrained and the unconstrained problems, We obtain, as a particular case of results, the existence of Nash equilibria for discounted constrained and unconstrained game models.

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

7.3.5 Optimal stochastic control and application on passive trajectography

The objective of this thesis [26] is to study optimal control problems in a partially observed discrete-time framework. In particular, we are interested in the stochastic optimal control problem for estimating the position of an underwater target from noisy measurements coming from a sensor owned by an underwater observer. We propose to use the Markov Decision Process (MDP) formalism, which is a family of controlled processes particularly well suited to modelling sequential stochastic optimisation problems. From a theoretical point of view, PDMs have been studied very intensively in the literature. There are numerous techniques for establishing theoretical results showing the existence of optimal strategies or characterising the properties of the value function. These techniques are centred around dynamic programming and linear programming. However, it should be emphasised that such approaches do not provide either explicit solutions (except for very specific cases) or approximations of these optimal solutions, particularly in the context of partial and noisy observations. This thesis will focus on developing new theoretical/numerical approaches to approximation techniques for solving partially observed PDMs, analysing the problem of localisation and tracking in the plane and in a context of angular observations, analysing the convergence of the proposed approximation schemes and implementing an algorithm for solving the problem described at the beginning of this paragraph.

Participants: Romain Namyst (ASTRAL), François Dufour (ASTRAL), Alexandre Genadot (ASTRAL).

7.4 Signal processing, artifical evolution and neural networks

7.4.1 Benefits of Zero-Phase or Linear Phase Filters to Design Multiscale Entropy: Theory and Application

In various applications, multiscale entropy (MSE) is often used as a feature to characterize the complexity of the signals in order to classify them. It consists of estimating the sample entropies (SEs) of the signal under study and its coarse-grained (CG) versions, where the CG process amounts to (1) filtering the signal with an average filter whose order is the scale and (2) decimating the filter output by a factor equal to the scale. In [19], we propose to derive a new variant of the MSE. Its novelty stands in the way to get the sequences at different scales by avoiding distortions during the decimation step. To this end, a linear-phase or null-phase low-pass filter whose cutoff frequency is well suited to the scale is used. Interpretations on how the MSE behaves and illustrations with a sum of sinusoids, as well as white and pink noises, are given. Then, an application to detect attentional tunneling is presented. It shows the benefit of the new approach in terms of p value when one aims at differentiating the set of MSEs obtained in the attentional tunneling state from the set of MSEs obtained in the nominal state. It should be noted that CG versions can be replaced not only for the MSE but also for other variants.

Participants: Eric Grivel, Bastien Berthelot, Gaetan Colin, Pierrick Legrand (ASTRAL), Vincent Ibanez.

7.4.2 Reservoir Computing for Short High-Dimensional Time Series: an Application to SARS-CoV-2 Hospitalization Forecast

In [24], we aimed at forecasting the number of SARS-CoV-2 hospitalized patients at 14 days to help anticipate the bed requirements of a large scale hospital using public data and electronic health records data. Previous attempts led to mitigated performance in this high-dimension setting; we introduce a novel approach to time series forecasting by providing an alternative to conventional methods to deal with high number of potential features of interest (409 predictors). We integrate Reservoir Computing (RC) with feature selection using a genetic algorithm (GA) to gather optimal non-linear combinations of inputs to improve prediction in sample-efficient context. We illustrate that the RC-GA combination exhibits excellent performance in forecasting SARS-CoV-2 hospitalizations. This approach outperformed the use of RC alone and other conventional methods: LSTM, Transformers, Elastic-Net, XGBoost. Notably, this work marks the pioneering use of RC (along with GA) in the realm of short and high-dimensional time series, positioning it as a competitive and innovative approach in comparison to standard methods.

This work is completed in [23]. This study focuses on forecasting SARS-CoV-2 hospitalizations 14 days ahead at Bordeaux University Hospital using a high-dimensional dataset with 409 predictors and 586 observations, combining public data and electronic health records. Previous research showed that integrating reservoir computing (RC) with genetic algorithm (GA) for hyperparameter optimization and feature selection outperformed state-of-the-art methods. However, the behavior of RC-GA under high-dimensional conditions is not well understood. This work examines the impact of GA hyperparameters (GA-HP), specifically the mutation probability of feature selection and the extent of mutation on a critical RC hyperparameter (RC-HP), the leaking rate. GA-HP significantly influence feature selection and RC-HP optimization. Higher mutation rates led to a more diverse set of selected features and fewer total features, resulting in increased mean absolute error (MAE) on the training set but comparable MAE on the test set, suggesting reduced overfitting. Conversely, lower mutation rates of categorical genes and leaking rates correlated with slightly poorer performance, indicating potential lack of exploration during RC-HP selection. Notably, a bimodal Ferté et al. convergence of the leaking rate was observed, with lower leaking rates enhancing training performance but slightly diminishing test performance. The RC-GA approach demonstrated robust behavior with higher leaking rates, possibly due to increased regularization effects from higher ridge values. Optimizing RC-GA in a high-dimensional setting remains challenging. Current practices using the median forecast of the top 40 RC-HP sets might not be optimal. Enhancing GA for high-dimensional contexts and exploring complementary RC-HP sets instead of selecting the top 40 might further improve forecasting accuracy.

Participants: Thomas Ferte, Dan Dutartre, Boris Heijblum, Romain Griffier, Vianney Jouhet, Rodolphe Thiébaut, Pierrick Legrand (ASTRAL), Xavier Hinaut.

7.4.3 Real-time simulation of nonlinear audio effects using artificial intelligence

Certain products in the realm of music technology have uniquely desirable sonic characteristics that are often sought after by musicians. These characteristics are often due to the nonlinearities of their electronic circuits. We are concerned with preserving the sound of this gear through digital simulations and making them widely available to numerous musicians. This field of study has seen a large rise in the use of neural networks for the simulation in recent years. In [27], we apply neural networks for the task. Particularly, we focus on real-time capable black-box methods for nonlinear effects modelling, with the guitarist in mind. We cover the current state-of-the-art and identify areas warranting improvement or study with a final goal of product development. A first step of identifying architectures capable of real-time processing in a streaming manner is followed by augmenting and improving these architectures and their training pipeline through a number of methods. These methods include continuous integration with unit testing, automatic hyperparameter optimisation, and the use of transfer learning. A real-time prototype utilising a custom C++ backend is created using these methods. A study in real-time anti-aliasing for black-box models is presented as it was found that these networks exhibit high amounts of aliasing distortion. Work on user control incorporation is also started for a comprehensive simulation of the analogue systems. This enables a full range of tone-shaping possibilities for the end user. The performance of the approaches presented is assessed both through objective and subjective evaluation. Finally, a number of possible directions for future work are also presented.

Participants: Tara Vanhatalo (ASTRAL), Pierrick Legrand (ASTRAL), Myriam Desainte-Catherine.

7.4.4 Tackling Long-Range Dependencies in Dynamic Range Compression Modeling via Deep Learning

Deep learning has increasingly been studied as a black-box method for audio effects modeling. While successful for some effects like equalizers or guitar amplifiers, learning long-range dependencies remains challenging. In [25], we discuss the disparity between streamed target and windowed target, and introduce the state prediction network, a general method addressing long-range dependencies in stateful models. We propose a novel model architecture, conduct hyperparameter search, and compare it with stateless and recurrent state-of-the-art models. Our architecture achieves better spectral accuracy with reduced training costs.

Participants: Yann Bourdin (ASTRAL), Pierrick Legrand (ASTRAL), Fanny Roche.

8 Bilateral contracts and grants with industry

8.1 Bilateral contracts with industry

Naval Group

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

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.

8.2 Bilateral Grants with Industry

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

Orosys

Participants: Tara Vanhatalo, Pierrick Legrand.

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

Arturia

Participants: Yann Bourdin, Pierrick Legrand.

Within the framework of Yann Bourdin's Cifre PhD thesis on the stochastic modeling of audio compressor and nonlinear audio effects, a strong collaboration was established between the company Arturia and the ASTRAL team.

9 Partnerships and cooperations

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

9.1 International initiatives

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

9.1.1 Participation in other International Programs

Scientific cooperation with Spain funded by the Spanish Ministry of Science and Innovation (reference number PID2021-122442NB-I00). This project focuses on the analysis and control of deterministic/stochastic dynamic systems and on game theory (2022-2025).

9.2 International research visitors

9.2.1 Visits of international scientists

Other international visits to the team

Professor Leonardo Trujillo

Status researcher

ITT Tijuana

Country: Mexique

Dates: Octobre 2024

Context of the visit: EA conference

Mobility program/type of mobility: research stay

Professor Oswaldo Costa

Status researcher

Escola Politécnica da Universidade São-Paulo

Country : Brasil

Context of the visit: research stay

Professor Tomàs Prieto-Rumeau

Status researcher

Facultad de Ciencias, UNED, Madrid

Country : Spain

Context of the visit: research stay

9.3 European initiatives

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

9.3.1 Other european programs/initiatives

Un réponse à l'appel à projet mobilité de Bdx-INP remporté par Pierrick Legrand et Jérôme Saracco. Mission de 5000 euros au Mexique.

9.4 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.

10 Dissemination

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

10.1 Promoting scientific activities

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

Vulgarisation scientifique, table ronde à la Data Party organisée par Bordeaux Metropole et l'association des petits débrouillards.

10.1.1 Scientific events: organisation

Organisation de la conférence EA 2024.

General chair, scientific chair Pierrick Legrand a été general chair de la conférence EA 2024.

Member of the organizing committees Les membres de l'équipe ont fait partie de divers comités d'organisation de conférences et d'events.

10.1.2 Scientific events: selection

Chair of conference program committees Pierrick Legrand a été general chair de la conférence EA 2024.

Member of the conference program committees Comme chaque année, les membres de l'équipe ASTRAL ont fait partie de plusieurs comités de programmes.

Reviewer Comme chaque année, les membres de l'équipe ASTRAL ont reviewé des articles de revue ou de conférences internationales.

10.1.3 Journal

Member of the editorial boards :
François Dufour is

- Corresponding Editor pour la revue SIAM Journal on Control Optimization depuis avril 2018 (SIAM - SICON).
- Associate Editor pour la revue Applied Mathematics and Optimization depuis janvier 2018 (AMO).
- Associate Editor pour la revue Stochastics: An International Journal of Probability and Stochastic Processes depuis juillet 2018 (Stochastics).
- Associate Editor pour la revue Mathematical Control and Related Fields depuis Janvier 2023 (Math. Control Related Fields).

Pierrick Legrand is

- GPEM editorial board member
- EA, LNCS, SPRINGER editorial board member

Reviewer - reviewing activities Comme chaque année, les membres de l'équipe ASTRAL ont reviewé des articles de revue ou de conférences internationales.

10.1.4 Invited talks

Pierrick Legrand was invited as keynote speaker at the NEO 2024 conference.

10.1.5 Scientific expertise

Pierrick Legrand est consultant pour la société LexisNexis.

10.2 Teaching - Supervision - Juries

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

10.2.1 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.
- Alexandre Genadot is the head of the MIASHS Licence of the Université de Bordeaux.
- Pierrick Legrand is in charge of the IBM Chair at ENSC.
- Licence : P. Legrand, Espaces Euclidiens, 46,5h, L2, Université de Bordeaux, France.
- Licence : P. Legrand, Informatique pour les mathématiques, 30h, L3, Université de Bordeaux, France.
- DU : P. Legrand, Evolution Artificielle, Big data, 8h, DU, Bordeaux INP, France.
- Engineer School: Signal processing, ENSC, Bordeaux, 1A, France.
- Engineer School: Signal processing, 54 hours, ENSC, Bordeaux, 2A, 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

10.2.2 Supervision

- Les membres de l'équipe ont participé à divers encadrements de stages de master, de PFE, de projets informatique individuels, etc.
- Les membres de l'équipe encadrent des thèses de doctorat.

10.2.3 Juries

- Présidence de jury du baccalaureat : P. Legrand
- Jury de licence MIASHS et de master Sciences Cognitives : M. Chavent, A. Genadot, P. Legrand
- Jury d'année en école d'ingénieur (ENSC) : J. Sarraco, P. Legrand.

10.3 Popularization

Participants: Francois Dufour, Pierre del Moral, Alexandre Genadot, Marie Chavent, Jérôme Saracco, Enzo Iglesis, Denis Arrivault, Pierrick Legrand.

10.3.1 Participation in Live events

Table ronde autour de l'IA et des data à la data party.

11 Scientific production

11.1 Major publications

[1] 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.

[2] 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.

[3] 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.

[4] A. Genadot. 'Contributions à l'étude des processus markoviens déterministes par morceaux et de décision ainsi qu'à l'étude de l'enquête Bourciez'. Université de Bordeaux, 25th Oct. 2023. URL: https://hal.science/tel-04326411.

[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] P. Legrand, A. Liefooghe, E. Keedwell, J. Lepagnot, L. Idoumghar, N. Monmarché and E. Lutton. *Artificial Evolution, 15th International Conference, Evolution Artificielle, EA 2022, Exeter, UK, October 31 - November 2, 2022, Revised Selected Papers*. Vol. 14091. Lecture Notes in Computer Science. Springer Nature Switzerland, 10th Sept. 2023. DOI: 10.1007/978-3-031-42616-2. URL: https://hal.science/hal-04194174.

[9] 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.

[10] 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.

[11] 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.

11.2 Publications of the year

International journals

[12] J. Albechaalany, M.-P. Ellies-Oury, J. Saracco, M. Campo, I. Richardson, P. Ertbjerg, S. Failla, B. Panea, J. Williams, M. Christensen and J.-F. Hocquette. 'Modelling the physiological, muscular, and sensory characteristics in relation to beef quality from 15 cattle breeds'. In: *Livestock Science* 280 (Feb. 2024), p. 105395. DOI: 10.1016/j.livsci.2023.105395. URL: https://hal.inrae.fr/hal-04417516 (cit. on p. 16).

[13] A. Aleksian, P. del Moral, A. Kurtzmann and J. Tugaut. 'Self-interacting diffusions: long-time behaviour and exit-problem in the convex case'. In: *ESAIM: Probability and Statistics* 28 (22nd Feb. 2024), pp. 46–61. DOI: 10.1051/ps/2023020. URL: https://hal.science/hal-01901145 (cit. on p. 18).

[14] O. Costa, F. Dufour and A. Genadot. 'Adaptive average control for piecewise deterministic Markov processes'. In: *Systems and Control Letters* 192 (Oct. 2024), p. 105894. DOI: 10.1016/j.sysconle.2024.105894. URL: https://hal.science/hal-04876309 (cit. on p. 18).

[15] O. L. D. V. Costa, F. Dufour and A. Genadot. 'Minimum Contrast Estimators for Piecewise Deterministic Markov Processes'. In: *Numerical Algebra, Control and Optimization* 15.1 (2025), pp. 1–14. DOI: 10.3934/naco.2024023. URL: https://hal.science/hal-04337828 (cit. on p. 18).

[16] F. Dufour. 'Absorbing Markov decision processes and their occupation measures'. In: *SIAM Journal on Control and Optimization* (2025). URL: https://hal.science/hal-04876371. In press (cit. on p. 19).

[17] F. Dufour and T. Prieto-Rumeau. 'Absorbing Markov Decision Processes'. In: *ESAIM: Control, Optimisation and Calculus of Variations* 30 (2024). DOI: 10.1051/cocv/2024002. URL: https://inria.hal.science/hal-04377071 (cit. on p. 19).

[18] F. Dufour and T. Prieto-Rumeau. 'Nash equilibria for total expected reward absorbing Markov games: the constrained and unconstrained cases'. In: *Applied Mathematics and Optimization* 89.2 (2024), p. 34. DOI: 10.1007/s00245-023-10095-1. URL: https://inria.hal.science/hal-04377070 (cit. on p. 19).

[19] E. Grivel, B. Berthelot, G. Colin, P. Legrand and V. Ibanez. 'Benefits of Zero-Phase or Linear Phase Filters to Design Multiscale Entropy: Theory and Application'. In: *Entropy* 26(4).332 (14th Apr. 2024), pp. 1–27. DOI: 10.3390/e26040332. URL: https://hal.science/hal-04570432 (cit. on p. 20).

[20] P. del Moral, S. Hu, A. Jasra, H. Ruzayqat and X. Wang. 'Bayesian Parameter Inference for Partially Observed Diffusions using Multilevel Stochastic Runge-Kutta Methods'. In: *International Journal for Uncertainty Quantification* (2024). DOI: 10.1615/Int.J.UncertaintyQuantification.2024051131. URL: https://inria.hal.science/hal-04877965 (cit. on p. 17).

[21] P. del Moral and X. Wang. 'Geometric Convergence and Concentration Inequalities for the Feynman–Kac Genetic Algorithm'. In: *Frontiers of Mathematics in China* 19.2 (5th Mar. 2024), pp. 321–334. DOI: 10.1007/s11464-022-0089-z. URL: https://inria.hal.science/hal-04877959 (cit. on p. 18).

[22] P. Neuvial, N. Randriamihamison, M. Chavent, S. Foissac and N. Vialaneix. 'A two-sample tree-based test for hierarchically organized genomic signals'. In: *Journal of the Royal Statistical Society: Series C Applied Statistics* (14th Mar. 2024), qlae011. DOI: 10.1093/jrsssc/qlae011. URL: https://hal.inrae.fr/hal-04516167 (cit. on p. 17).

International peer-reviewed conferences

[23] T. Ferté, D. Dutartre, B. P. Hejblum, R. Griffier, V. Jouhet, R. Thiébaut, X. Hinaut and P. Legrand. 'Optimizing Reservoir Computing with Genetic Algorithm for High-Dimensional SARS-CoV-2 Hospitalization Forecasting: Impacts of Genetic Algorithm Hyperparameters on Feature Selection and Reservoir Computing Hyperparameter Tuning'. In: 16th International Conference, Évolution Artificielle, EA 2024. Lecture Notes in Computer Science. Bordeaux, France, 29th Oct. 2024. URL: https://inria.hal.science/hal-04905975 (cit. on p. 20).

Conferences without proceedings

[24] T. Ferté, D. Dutartre, B. P. Hejblum, R. Griffier, V. Jouhet, R. Thiébaut, P. Legrand and X. Hinaut. 'Reservoir Computing for Short High-Dimensional Time Series: an Application to SARS-CoV-2 Hospitalization Forecast'. In: ICML'24: Proceedings of the 41st International Conference on Machine Learning. Vol. 235. Proceedings of Machine Learning Research. Vienna, Austria, 8th July 2024, pp. 13570–13591. DOI: 10.5555/3692070. URL: https://hal.science/hal-04693930 (cit. on p. 20).

Edition (books, proceedings, special issue of a journal)

[25] *Tackling Long-Range Dependencies in Dynamic Range Compression Modeling via Deep Learning*. Evolution Artificielle. Oct. 2024. URL: https://inria.hal.science/hal-04907223 (cit. on p. 21).

Doctoral dissertations and habilitation theses

[26] R. Namyst. 'Optimal stochastic control and application on passive trajectography'. Université de Bordeaux, 6th Dec. 2024. URL: https://theses.hal.science/tel-04889756 (cit. on p. 19).

[27] T. Vanhatalo. 'Real-time simulation of nonlinear audio effects using artificial intelligence'. Université de Bordeaux, 29th Apr. 2024. URL: https://inria.hal.science/tel-04921220 (cit. on p. 21).

Reports & preprints

[28] O. D. Akyildiz, P. del Moral and J. Miguez. *Gaussian entropic optimal transport: Schrödinger bridges and the Sinkhorn algorithm*. 24th Dec. 2024. URL: https://inria.hal.science/hal-04877986.

[29] M. Caffarel, P. del Moral and L. de Montella. *On the Mathematical foundations of Diffusion Monte Carlo*. 22nd Jan. 2024. URL: https://hal.science/hal-04409602.

[30] M. Chavent and G. Chavent. *From explained variance of correlated components to PCA without orthogonality constraints*. 6th Feb. 2024. URL: https://inria.hal.science/hal-04442489.

[31] F. Dufour, A. Genadot and R. Namyst. *The Bearing Only Localization Problem via Partially Observed Markov Decision Process*. 6th May 2024. URL: https://hal.science/hal-04568986.

[32] A. Genadot, E. Iglésis and L. de Montella. *About Fisher Information Matrix and Self-Optimizing Control in Bearing Only Localization*. 17th Jan. 2025. URL: https://inria.hal.science/hal-04894826.

[33] E. Iglésis and L. de Montella. *Target Motion Analysis Using Combined Bearings and Underwater Sound Pressure Levels*. 14th Nov. 2024. URL: https://inria.hal.science/hal-04796331.

11.3 Cited publications

[34] E. Altman. *Constrained Markov decision processes*. Stochastic Modeling. Chapman & Hall/CRC, Boca Raton, FL, 1999, pp. xii+242 (cit. on p. 12).

[35] 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 (cit. on p. 12).

[36] 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 (cit. on p. 12).

[37] 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 (cit. on p. 12).

[38] 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 (cit. on p. 12).

[39] P. Del Moral. *Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer-Verlag, New York, 2004, p. 573 (cit. on pp. 10, 11).

[40] 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 (cit. on pp. 10, 11).

[41] 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 (cit. on p. 10).

[42] 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 (cit. on p. 10).

[43] 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 (cit. on pp. 10, 11).

[44] 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 (cit. on pp. 10, 11).

[45] 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 (cit. on pp. 10, 11).

[46] P. Del Moral and S. Penev. *Stochastic Processes: From Applications to Theory*. Chapman and Hall/CRC, 2017 (cit. on pp. 10, 11).

[47] 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 (cit. on p. 10).

[48] P. Del Moral. 'Non Linear Filtering: Interacting Particle Solution'. In: Markov Processes and Related Fields 2.4 (1996), pp. 555–580 (cit. on pp. 10, 11).

[49] E. Dynkin and A. Yushkevich. *Controlled Markov processes*. Vol. 235. Grundlehren der Mathematischen Wissenschaften. Berlin: Springer-Verlag, 1979, pp. xvii+289 (cit. on p. 12).

[50] J. Filar and K. Vrieze. *Competitive Markov decision processes*. New York: Springer-Verlag, 1997, pp. xii+393 (cit. on p. 12).

[51] 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 (cit. on p. 12).

[52] O. Hernández-Lerma. *Adaptive Markov control processes*. Vol. 79. Applied Mathematical Sciences. New York: Springer-Verlag, 1989, pp. xiv+148 (cit. on p. 12).

[53] 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 (cit. on p. 12).

[54] 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 (cit. on p. 12).

[55] 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 (cit. on p. 12).

[56] 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 (cit. on p. 12).

[57] 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 (cit. on p. 12).

[58] V. kolokoltsov. 'Nonlinear Markov Processes and Kinetic Equations'. In: Cambridge Univ. Press (2010) (cit. on p. 10).

[59] A. B. Piunovskiy. *Examples in Markov decision processes*. Vol. 2. Imperial College Press Optimization Series. Imperial College Press, London, 2013, pp. xiv+293 (cit. on p. 12).

[60] 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 (cit. on p. 12).

[61] 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 (cit. on p. 12).

[62] 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 (cit. on p. 12).