

RESEARCH CENTRE

Rennes - Bretagne Atlantique

IN PARTNERSHIP WITH:

Université Rennes 1, CNRS

2021

ACTIVITY REPORT

Project-Team

SIMSMART

## **SIMulating Stochastic Models with pARTicles**

IN COLLABORATION WITH: Institut de recherche mathématique de  
Rennes (IRMAR)

**DOMAIN**

**Applied Mathematics, Computation and  
Simulation**

**THEME**

**Stochastic approaches**

# Contents

<b>Project-Team SIMSMART</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>2</b>
<b>3 Research program</b>	<b>3</b>
<b>4 Application domains</b>	<b>4</b>
4.1 Domain 1 – Computational Physics . . . . .	4
4.2 Domain 2 – Meteorology . . . . .	5
4.3 Other Applicative Domains . . . . .	5
<b>5 Social and environmental responsibility</b>	<b>5</b>
5.1 Footprint of research activities . . . . .	5
<b>6 New software and platforms</b>	<b>5</b>
6.1 New software . . . . .	5
6.1.1 Screening4L0Problem . . . . .	5
6.1.2 Screen&Relax . . . . .	6
6.1.3 npSEM . . . . .	6
6.1.4 NHMSAR . . . . .	6
6.1.5 3D Winds Fields Profiles . . . . .	6
6.1.6 Screening4SLOPE . . . . .	7
<b>7 New results</b>	<b>7</b>
7.1 Objective 1 – Rare events and Monte Carlo simulation . . . . .	7
7.2 Objective 2 – New topics in particle filtering . . . . .	7
7.3 Objective 3 – Semi-parametric statistics . . . . .	7
7.4 Objective 4 – Model Reduction and Sparsity . . . . .	8
<b>8 Bilateral contracts and grants with industry</b>	<b>8</b>
8.1 Bilateral contracts with industry . . . . .	8
8.2 Preliminary collaboration . . . . .	9
<b>9 Partnerships and cooperations</b>	<b>9</b>
9.1 National initiatives . . . . .	9
9.1.1 ANR . . . . .	9
<b>10 Dissemination</b>	<b>9</b>
10.1 Promoting scientific activities . . . . .	9
10.1.1 Invited talks . . . . .	9
10.2 Teaching - Supervision - Juries . . . . .	9
10.2.1 Teaching . . . . .	9
10.2.2 Juries . . . . .	10
10.2.3 Supervision . . . . .	10
<b>11 Scientific production</b>	<b>11</b>
11.1 Publications of the year . . . . .	11

## Project-Team SIMSMART

*Creation of the Project-Team: 2019 January 01*

### Keywords

#### Computer sciences and digital sciences

- A6. – Modeling, simulation and control
  - A6.1. – Methods in mathematical modeling
    - A6.1.1. – Continuous Modeling (PDE, ODE)
    - A6.1.2. – Stochastic Modeling
    - A6.1.4. – Multiscale modeling
  - A6.2. – Scientific computing, Numerical Analysis & Optimization
    - A6.2.1. – Numerical analysis of PDE and ODE
    - A6.2.2. – Numerical probability
    - A6.2.3. – Probabilistic methods
    - A6.2.4. – Statistical methods
    - A6.2.5. – Numerical Linear Algebra
    - A6.2.6. – Optimization
  - A6.3. – Computation-data interaction
    - A6.3.1. – Inverse problems
    - A6.3.2. – Data assimilation
    - A6.3.4. – Model reduction
    - A6.3.5. – Uncertainty Quantification
  - A6.5. – Mathematical modeling for physical sciences
    - A6.5.2. – Fluid mechanics
    - A6.5.3. – Transport
    - A6.5.5. – Chemistry

#### Other research topics and application domains

- B1. – Life sciences
- B2. – Health
- B3. – Environment and planet
  - B3.2. – Climate and meteorology
- B4. – Energy
  - B4.2. – Nuclear Energy Production
    - B4.2.1. – Fission
- B5.3. – Nanotechnology
- B5.5. – Materials

## 1 Team members, visitors, external collaborators

### Research Scientists

- Mathias Rousset [Team leader, Inria, Researcher, HDR]
- Frédéric C  rou [Inria, Researcher]
- C  dric Herzet [Inria, Researcher]
- Patrick H  as [Inria, Researcher]
- Fran  ois Le Gland [Inria, Senior Researcher]

### Faculty Member

- Val  rie Monbet [Univ de Rennes I, Professor, HDR]

### PhD Students

- Francois Ernoult [Univ de Rennes I]
- Theo Guyard [INSA Rennes, from Oct 2021]
- Said Obakrim [Univ de Rennes I]
- Thu Le Tran [Univ de Rennes I]

### Interns and Apprentices

- Jules Berry [Inria, from Jun 2021 until Jul 2021]
- Hala Bouzidi [Inria, from May 2021 until Jul 2021]

### Administrative Assistants

- Fabienne Cuyollaa [Inria, until Jun 2021]
- Gunther Tessier [Inria, from Sep 2021]

## 2 Overall objectives

As the constant surge of computational power is nurturing scientists into simulating the most detailed features of reality, from complex molecular systems to climate or weather forecast, the computer simulation of physical systems is becoming reliant on highly complex stochastic dynamical models and very abundant observational data. The complexity of such models and of the associated observational data stems from intrinsic physical features, which do include high dimensionality as well as intricate temporal and spatial multi-scales. It also results in much less control over simulation uncertainty.

Within this highly challenging context, SIMSMART positions itself as a mathematical and computational probability and statistics research team, dedicated to *Monte Carlo simulation* methods. Such methods include in particular particle Monte Carlo methods for rare event simulation, data assimilation and model reduction, with application to stochastic random dynamical physical models. The main objective of SIMSMART is to disrupt this now classical field by creating deeper mathematical frameworks adapted to the management of contemporary highly sophisticated physical models.

### 3 Research program

**Introduction.** Computer simulation of physical systems is becoming increasingly reliant on highly complex models, as the constant surge of computational power is nurturing scientists into simulating the most detailed features of reality – from complex molecular systems to climate/weather forecast.

Yet, when modeling physical reality, bottom-up approaches are stumbling over intrinsic difficulties. First, the timescale separation between the fastest simulated microscopic features, and the macroscopic effective slow behavior becomes huge, implying that the fully detailed and direct long time simulation of many interesting systems (*e.g.* large molecular systems) are out of reasonable computational reach. Second, the chaotic dynamical behaviors of the systems at stake, coupled with such multi-scale structures, exacerbate the intricate uncertainty of outcomes, which become highly dependent on intrinsic chaos, uncontrolled modeling, as well as numerical discretization. Finally, the massive increase of observational data addresses new challenges to classical data assimilation, such as dealing with high dimensional observations and/or extremely long time series of observations.

**SIMSMART Identity.** Within this highly challenging applicative context, SIMSMART positions itself as a computational probability and statistics research team, with a mathematical perspective. Our approach is based on the use of *stochastic modeling* of complex physical systems, and on the use of *Monte Carlo simulation* methods, with a strong emphasis on dynamical models. The two main numerical tasks of interest to SIMSMART are the following: (i) simulating with pseudo-random number generators - a.k.a. *sampling* - dynamical models of random physical systems, (ii) sampling such random physical dynamical models given some real observations - a.k.a. *Bayesian data assimilation*. SIMSMART aims at providing an appropriate mathematical level of abstraction and generalization to a wide variety of Monte Carlo simulation algorithms in order to propose non-superficial answers to both *methodological and mathematical* challenges. The issues to be resolved include computational complexity reduction, statistical variance reduction, and uncertainty quantification.

**SIMSMART Objectives.** The main objective of SIMSMART is to disrupt this now classical field of particle Monte Carlo simulation by creating deeper mathematical frameworks adapted to the challenging world of complex (*e.g.* high dimensional and/or multi-scale), and massively observed systems, as described in the beginning of this introduction.

To be more specific, we will classify SIMSMART objectives using the following four intertwined topics:

- Objective 1: Rare events and random simulation.
- Objective 2: High dimensional and advanced particle filtering.
- Objective 3: Non-parametric approaches.
- Objective 4: Model reduction and sparsity.

Rare events Objective 1 are ubiquitous in random simulation, either to accelerate the occurrence of physically relevant random slow phenomena, or to estimate the effect of uncertain variables. Objective 1 will be mainly concerned with particle methods where *splitting* is used to enforce the occurrence of rare events.

The problem of high dimensional observations, the main topic in Objective 2, is a known bottleneck in filtering, especially in non-linear particle filtering, where linear data assimilation methods remain the state-of-the-art approaches.

The increasing size of recorded observational data and the increasing complexity of models also suggest to devote more effort into non-parametric data assimilation methods, the main issue of Objective 3.

In some contexts, for instance when one wants to compare solutions of a complex (*e.g.* high dimensional) dynamical systems depending on uncertain parameters, the construction of relevant reduced-order models becomes a key topic. Model reduction aims at proposing efficient algorithmic procedures for the resolution (to some reasonable accuracy) of high-dimensional systems of parametric equations.

This overall objective entails many different subtasks: 1) the identification of low-dimensional surrogates of the target “solution” manifold, 2) The devise of efficient methodologies of resolution exploiting low-dimensional surrogates, 3) The theoretical validation of the accuracy achievable by the proposed procedures. This is the content of Objective 4.

With respect to volume of research activity, Objective 1, Objective 4 and the sum (Objective 2+Objective 3) are comparable.

Some new challenges in the simulation and data assimilation of random physical dynamical systems have become prominent in the last decade. A first issue (i) consists in the intertwined problems of simulating on large, macroscopic random times, and simulating *rare events* (see Objective 1). The link between both aspects stems from the fact that many effective, large times dynamics can be approximated by sequences of rare events. A second, obvious, issue (ii) consists in managing *very abundant observational data* (see Objective 2 and 3). A third issue (iii) consists in quantifying *uncertainty/sensitivity/variance* of outcomes with respect to models or noise. A fourth issue (iv) consists in managing *high dimensionality*, either when dealing with complex prior physical models, or with very large data sets. The related increase of complexity also requires, as a fifth issue (v), the construction of *reduced models* to speed-up comparative simulations (see Objective 4). In a context of very abundant data, this may be replaced by a sixth issue (vi) where complexity constraints on modeling is replaced by the use of *non-parametric statistical inference* (see Objective 3).

Hindsight suggests that all the latter challenges are related. Indeed, the contemporary digital condition, made of a massive increase in computational power and in available data, is resulting in a demand for more complex and uncertain models, for more extreme regimes, and for using inductive approaches relying on abundant data. In particular, uncertainty quantification (item (iii)) and high dimensionality (item (iv)) are in fact present in all 4 Objectives considered in SimSmart.

## 4 Application domains

### 4.1 Domain 1 – Computational Physics

The development of large-scale computing facilities has enabled simulations of systems at the *atomistic scale* on a daily basis. The aim of these simulations is to bridge the time and space scales between the macroscopic properties of matter and the stochastic atomistic description. Typically, such simulations are based on the ordinary differential equations of classical mechanics supplemented with a random perturbation modeling temperature, or collisions between particles.

Let us give a few examples. In bio-chemistry, such simulations are key to predict the influence of a ligand on the behavior of a protein, with applications to drug design. The computer can thus be used as a *numerical microscope* in order to access data that would be very difficult and costly to obtain experimentally. In that case, a rare event (Objective 1) is given by a macroscopic system change such as a conformation change of the protein. In nuclear safety, such simulations are key to predict the transport of neutrons in nuclear plants, with application to assessing aging of concrete. In that case, a rare event is given by a high energy neutron impacting concrete containment structures.

A typical model used in molecular dynamics simulation of open systems at given temperature is a stochastic differential equation of Langevin type. The large time behavior of such systems is typically characterized by a hopping dynamics between ‘metastable’ configurations, usually defined by local minima of a potential energy. In order to bridge the time and space scales between the atomistic level and the macroscopic level, specific algorithms enforcing the realization of rare events have been developed. For instance, splitting particle methods (Objective 1) have become popular within the computational physics community only within the last few years, partially as a consequence of interactions between physicists and Inria mathematicians in ASPI (parent of SIMSMART) and MATERIALS project-teams.

SIMSMART also focuses on various models described by partial differential equations (reaction-diffusion, conservation laws), with unknown parameters modeled by random variables.

## 4.2 Domain 2 – Meteorology

The traditional trend in data assimilation in geophysical sciences (climate, meteorology) is to use as prior information some very complex deterministic models formulated in terms of fluid dynamics and reflecting as much as possible the underlying physical phenomenon (see *e.g.*). Weather/climate forecasting can then be recast in terms of a Bayesian filtering problem (see Objective 2) using weather observations collected *in situ*.

The main issue is therefore to perform such Bayesian estimations with very expensive infinite dimensional prior models, and observations in large dimension. The use of some linear assumption in prior models (Kalman filtering) to filter non-linear hydrodynamical phenomena is the state-of-the-art approach, and a current field of research, but is plagued with intractable instabilities.

This context motivates two research trends: (i) the introduction of non-parametric, model-free prior dynamics constructed from a large amount of past, recorded real weather data; and (ii) the development of appropriate non-linear filtering approaches (Objective 2 and Objective 3).

SIMSMART will also test its new methods on multi-source data collected in North-Atlantic paying particular attention to coastal areas (*e.g.* within the inter-Labex SEACS).

## 4.3 Other Applicative Domains

SIMSMART focuses on various applications including:

- Tracking and hidden Markov models.
- Robustness and certification in Machine Learning.

# 5 Social and environmental responsibility

## 5.1 Footprint of research activities

Zero: no travel by plane; no intensive numerical computing.

# 6 New software and platforms

The following softwares have been released or updated.

## 6.1 New software

### 6.1.1 Screening4L0Problem

**Keywords:** Global optimization, Sparsity

**Functional Description:** This software contains "Branch and bound" optimization routines exploiting "screening" acceleration rules for solving sparse representation problems involving the L0 pseudo-norm.

**URL:** <https://gitlab.insa-rennes.fr/Theo.Guyard/bnb-screening>

**Publication:** hal-03462171

**Contact:** Cedric Herzet

**Participants:** Clement Elvira, Theo Guyard

### 6.1.2 Screen&Relax

**Keywords:** Optimization, Sparsity

**Functional Description:** This software provides optimization routines to efficiently solve the "Elastic-Net" problem.

**URL:** <https://gitlab.insa-rennes.fr/Theo.Guyard/screen-and-relax>

**Publication:** hal-03462191

**Contact:** Cedric Herzet

**Participants:** Clement Elvira, Theo Guyard

### 6.1.3 npSEM

**Name:** Stochastic expectation-maximization algorithm for non-parametric state-space models

**Keyword:** Statistic analysis

**Functional Description:** npSEM is the combination of a non-parametric estimate of the dynamic using local linear regression (LLR), a conditional particle smoother and a stochastic Expectation-Maximization (SEM) algorithm. Further details of its construction and implementation are introduced in the article An algorithm for non-parametric estimation in state-space models of authors "T.T.T. Chau, P. Ailliot, V. Monbet", <https://doi.org/10.1016/j.csda.2020.107062>.

**URL:** <https://github.com/tchau218/npSEM>

**Contact:** Thi Tuyet Trang Chau

### 6.1.4 NHMSAR

**Name:** Non-Homogeneous Markov Switching Autoregressive Models

**Keyword:** Statistical learning

**Functional Description:** Calibration, simulation, validation of (non-)homogeneous Markov switching autoregressive models with Gaussian or von Mises innovations. Penalization methods are implemented for Markov Switching Vector Autoregressive Models of order 1 only. Most functions of the package handle missing values.

**URL:** <https://CRAN.R-project.org/package=NHMSAR>

**Contact:** Valérie Monbet

### 6.1.5 3D Winds Fields Profiles

**Keywords:** 3D modeling, Optic-flow, Atmosphere

**Functional Description:** The algorithm computes 3D Atmospheric Motion Vectors (AMVs) vertical profiles, using incomplete maps of humidity, temperature and ozone concentration observed in a range of isobaric levels. The code is implemented for operational use with the Infrared Atmospheric Sounding Interferometer (IASI) carried on the MetOp satellite.

**URL:** [https://www.esa.int/Applications/Observing\\_the\\_Earth/Meteorological\\_missions/MetOp/About\\_IASI](https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/MetOp/About_IASI)

**Contact:** Patrick Heas



### 6.1.6 Screening4SLOPE

**Keyword:** Optimization

**Functional Description:** This software provides optimization routines to solve the SLOPE problem by exploiting "safe screening" reduction techniques.

**URL:** <https://gitlab-research.centralesupelec.fr/2020elvirac/slope-screening>

**Publication:** hal-03400322

**Contact:** Cedric Herzet

**Participant:** Clement Elvira

## 7 New results

### 7.1 Objective 1 – Rare events and Monte Carlo simulation

#### Monte-Carlo simulation

**Participants:** Frédéric Cérou, Patrick Héas, Mathias Rousset, François Ernoult.

In [7, 8], we quantify the robustness of a trained network to input uncertainties with a stochastic simulation and a statistical hypothesis test: the network is deemed as locally robust if the estimated probability of failure is lower than a critical level. The procedure is based on an Importance Splitting simulation generating samples of rare events. We derive theoretical guarantees that are non-asymptotic w.r.t. sample size. Experiments tackling large scale networks outline the efficiency of our method making a low number of calls to the network function.

### 7.2 Objective 2 – New topics in particle filtering

#### Multitarget tracking in track-before-detect context

**Participants:** Audrey Cuillery, François Le Gland, Valérie Monbet.

This work [6, 9] introduces a new class of particle filters, that include an auxiliary Markov transition in their design. This approach can be seen as an extension of the auxiliary particle filter. A prototypical situation where crossover could be useful is multitarget tracking. Indeed, it may happen that some targets in a multitarget particle are good proxies, but are not going to be selected just because the other targets in the same multitarget particle are bad proxies. This is unfair, and a better design would be to produce shuffled multitarget particles such that the particle for each different target can be replicated from a different multitarget particle.

### 7.3 Objective 3 – Semi-parametric statistics

**Participants:** Valérie Monbet, Cédric Herzet, Thu Le Tran, Said Obakrim.

MIR spectroscopy is becoming an increasingly important tool potentially useful for diagnosis purposes especially by studying body fluids. However, such changes can be difficult to capture if the structure of the data is not considered. Our objective in [3] was to improve MIR spectra analysis by using approximation of the spectra by B-splines at different specific resolutions and to combine these spectra representations with a machine learning model to predict hepatic steatosis from serum study.

## 7.4 Objective 4 – Model Reduction and Sparsity

**Participants:** Patrick Héas, Cédric Herzet, Théo Guyard.

In the context of model reduction, an issue is to find fast algorithms to project onto low-dimensional, sparse models.

[2] studies the linear approximation of high-dimensional dynamical systems using low-rank dynamic mode decomposition. Searching this approximation in a data-driven approach is formalized as attempting to solve a low-rank constrained optimization problem. This problem is non-convex, and state-of-the-art algorithms are all sub-optimal. This paper shows that there exists a closed-form solution, which is computed in polynomial time, and characterizes the  $\ell^2$ -norm of the optimal approximation error.

Another avenue of research has been the study of the sparse surrogate in the context of “continuous” dictionaries, where the elementary signals forming the decomposition catalog are functions of some parameters taking their values in some continuous domain. In this context, we contributed to the theoretical characterization of the performance of a well-known algorithmic procedure, namely “orthogonal matching pursuit” (OMP). More specifically, we proposed the first theoretical analysis of the behavior of OMP in the continuous setup, see [1].

In [4], we address the problem of approximating the atoms of a parametric dictionary, commonly encountered in the context of sparse representations in “continuous” dictionaries. We focus on the case of translation-invariant dictionaries. We derive necessary and sufficient conditions characterizing the existence of an “interpolating” and “translation-invariant” low-rank approximation.

[5] deals with the sensor placement problem for an array designed for source localization. When it involves the identification of a few sources, the compressed sensing framework is known to find directions effectively thanks to sparse approximation. The present contribution intends to provide an answer to the following question: given a set of observations, how should we make the next measurement to minimize (some form of) uncertainty on the localization of the sources?

In [10], we propose a methodology to accelerate the resolution of the so-called “Sorted L-One Penalized Estimation” (SLOPE) problem. Our method leverages the concept of “safe screening”, well-studied in the literature for *group-separable* sparsity-inducing norms, and aims at identifying the zeros in the solution of SLOPE.

In [11], we present a novel screening methodology to safely discard irrelevant nodes within a generic branch-and-bound (BnB) algorithm solving the  $\ell_0$ -penalized least-squares problem. Our contribution is a set of two simple tests to detect sets of feasible vectors that cannot yield optimal solutions. This allows to prune nodes of the BnB exploration tree, thus reducing the overall solution time.

In [12], we propose a procedure to accelerate the resolution of the well-known “Elastic-Net” problem. Our procedure is based on the (partial) identification of the solution support and the reformulation of the original problem into a problem of reduced dimension. The identification of the support leverages the novel concept of “safe relaxing” where one aims at identifying non-zero coefficients of the solution. It can be viewed as a dual approach to “safe screening” introduced in the last decade and allowing to reduce the problem dimension using the identification of zero coefficients of the solution. We show numerically that combining both methodologies in a “Screen And Relax” strategy enables to significantly improve the tradeoff between complexity and accuracy achievable by standard resolution techniques.

## 8 Bilateral contracts and grants with industry

### 8.1 Bilateral contracts with industry

**Participants:** Valérie Monbet, Patrick Héas.

- **Scalian**, through the CIFRE PhD project of Gabriel Jouan supervised by Valérie Monbet, dedicated to weather forecast corrections.
- **LIFY air**, through the CIFRE PhD project of Ezzo-Ridah Bleza supervised by Valérie Monbet on IA for multi-source pollen detection.

## 8.2 Preliminary collaboration

Patrick Héas is collaborating with the agency European Organization for the Exploitation of Meteorological Satellites (**EUMETSAT**) of Darmstadt. The transfer focuses on the estimation of atmospheric 3D winds from the future hyperspectral instruments.

## 9 Partnerships and cooperations

**Participants:** Cédric Herzet.

### 9.1 National initiatives

#### 9.1.1 ANR

**ANR Melody (2020-2024):** Bridging geophysics and Machine Learning for the modeling, simulation and reconstruction of Ocean Dynamics.

Cédric Herzet is part of the MELODY project. The MELODY project aims to bridge the physical model-driven paradigm underlying ocean/atmosphere science and AI paradigms with a view to developing geophysically-sound learning-based and data-driven representations of geophysical flows accounting for their key features (e.g., chaos, extremes, high-dimensionality).

## 10 Dissemination

**Participants:** Mathias Rousset, Valérie Monbet, François Le Gland, Cédric Herzet, Patrick Héas, Frédéric Cérou.

### 10.1 Promoting scientific activities

#### 10.1.1 Invited talks

Cédric Herzet (with Clément Elvira):

- *Safe Rules for the Identification of Zeros in the Solutions of the SLOPE Problem* at **Slope Conference**.

Mathias Rousset:

- *Rare Event Simulation of Neutron Trajectories* at **Bath University's Neutron Transport Days**
- *Rare Event Simulation with Adaptive Splitting* at **RESIM**

### 10.2 Teaching - Supervision - Juries

#### 10.2.1 Teaching

Cédric Herzet has given:

- INSA RENNES, 5ième année de l'option Génie Mathématique, cours de Parcimonie en traitement du signal et des images, 16h de cours magistraux + responsable du module
- Ensai RENNES, Master international « Smart Data », cours « Foundations of Smart Sensing », 15h de cours magistraux
- Ensai RENNES, Master international « Smart Data », cours « Advanced topics in Smart Sensing », 3h de cours magistraux
- ENS RENNES, Master 1, cours « Traitement du signal », 10h de cours magistraux
- Ensai RENNES, Master 2, suivi de projets, 15h

François Le Gland has given

- a 2nd year course on **introduction to stochastic differential equations**, at INSA (institut national des sciences appliquées) Rennes, within the cursus in applied mathematical,
- a 3rd year course on **Bayesian filtering and particle approximation**, at ENSTA (école nationale supérieure de techniques avancées), Palaiseau, within the statistics and control module,
- a 3rd year course on **linear and nonlinear filtering**, at ENSAI (école nationale de la statistique et de l'analyse de l'information), Ker Lann, within the statistical engineering track.

Mathias Rousset has given a 24h specialized course 'Large Deviations Theory' in Master 2 Fundamental Mathematics, Univ Rennes.

Patrick Héas has given a 28h class on algorithms and complexity at ESIR (Ecole Supérieure d'Ingénieurs de Rennes), first year.

### 10.2.2 Juries

François Le Gland has been a reviewer for the PhD thesis of Camille Palmier (université de Bordeaux, advisor: Pierre Del Moral).

Mathias Rousset has been the reviewer for the PhD thesis of Alessandro Andrea Barp (Imperial College), advisor: Mark Girolami.

### 10.2.3 Supervision

#### Internships

- Cédric Herzet has supervised one Master 1 (IRMAR) internship (Jules Berry). Subject: sparsity.
- Patrick Héas has supervised one Master 1 (ENSAE) internship (Hala Bouzidi). Subject: Langevin Markov Chain Monte Carlo in high dimensional Bayesian statistics.

#### PhDs

Cédric Herzet has supervised the PhDs of:

- Le Tran Thu, PhD, co-supervision with V. Monbet (Université de Rennes 1). UR1 funding. Subject: automatic diagnosis from spectral data using sparse representations in continuous dictionaries.
- Théo Guyard, PhD. INSA funding. Subject: optimization algorithms for sparse representations.

Mathias Rousset have supervised:

- François Ernout, PhD (co-supervised with Frédéric Cérou). Funding: UR1 and Région Bretagne. Subject: small noise asymptotics of rare events Monte Carlo simulation algorithms
- Karim Tit, PhD, (co-supervised with Teddy Furon). Funding: CIFRE Thalès. Subject: statistical robustness of machine learning classifiers.

V. Monbet has also supervised:

- Le Thu Tran, PhD, Univ Rennes, Diagnosis Learning from Scarce Data with Sparse Representations in Continuous Dictionaries. Funding: 1/2 UR1 + 1/2 ANR AI4SDA
- Gabriel Jouan, PhD, Univ Rennes, Scalian, granted by CIFRE - Scalian. Co-supervision with A. Cuzol (UBS) et G. Monnier (Scalian).
- Ezzo-Ridah Bleza, PhD, Univ Bretagne Sud, Janasense, granted by CIFRE- LIFY air. Co-supervision: PF Marteau (UBS)
- Said Obakrim, PhD, Univ Rennes, Ifremer (co-supervision: N. Raillard (Ifremer) et P. Ailliot (UBO)). Funding: 1/2 UR1 + 1/2 Ifremer.

## 11 Scientific production

### 11.1 Publications of the year

#### International journals

- [1] C. Elvira, R. Gribonval, C. Soussen and C. Herzet. ‘When does OMP achieve exact recovery with continuous dictionaries?’ In: *Applied and Computational Harmonic Analysis* 51 (Mar. 2021), p. 39. DOI: [10.1016/j.acha.2020.12.002](https://doi.org/10.1016/j.acha.2020.12.002). URL: <https://hal.inria.fr/hal-02099464>.
- [2] P. Héas and C. Herzet. ‘Low-Rank Dynamic Mode Decomposition: An Exact and Tractable Solution’. In: *Journal of Nonlinear Science* 32.1 (2021), article n°8. DOI: [10.1007/s00332-021-09770-w](https://doi.org/10.1007/s00332-021-09770-w). URL: <https://hal.inria.fr/hal-03468966>.
- [3] D. Martin, V. Monbet, O. Sire, M. Le Corvec and O. Loréal. ‘Multi-resolution B-splines data compression improves MIR spectroscopy-based Health diagnostic efficiency’. In: *Talanta Open* 4 (Dec. 2021), p. 100063. DOI: [10.1016/j.talo.2021.100063](https://doi.org/10.1016/j.talo.2021.100063). URL: <https://hal.archives-ouvertes.fr/hal-03322143>.

#### International peer-reviewed conferences

- [4] F. Champagnat and C. Herzet. ‘Interpolating and translation-invariant approximations of parametric dictionaries’. In: EUSIPCO 2020 - 28th European Signal Processing Conference. Amsterdam, Netherlands: IEEE, 18th Jan. 2021, pp. 2011–2015. DOI: [10.23919/Eusipco47968.2020.9287460](https://doi.org/10.23919/Eusipco47968.2020.9287460). URL: <https://hal.inria.fr/hal-03070383>.
- [5] M. Courcoux-Caro, C. Vanwynsberghe, C. Herzet and A. Baussard. ‘Sequential Sensor Placement using Bayesian Compressed Sensing for Source Localization’. In: EUSIPCO 2020 - 28th European Signal Processing Conference. Vol. 2021. European Signal Processing Conference. Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 241–245. DOI: [10.23919/Eusipco47968.2020.9287709](https://doi.org/10.23919/Eusipco47968.2020.9287709). URL: <https://hal.inria.fr/hal-03070390>.
- [6] A. Cuillery and F. Le Gland. ‘Particle filters with auxiliary Markov transition. Application to crossover and to multitarget tracking’. In: Fusion 2021 - 24th International Conference on Information Fusion. Rustenburg, South Africa: IEEE, 1st Nov. 2021. URL: <https://hal.inria.fr/hal-03467557>.
- [7] K. Tit, T. Furon and M. Rousset. ‘Efficient Statistical Assessment of Neural Network Corruption Robustness’. In: NeurIPS 2021 - 35th Conference on Neural Information Processing Systems. Vol. 34. Advances in Neural Information Processing Systems proceedings. Sydney (virtual), Australia, Dec. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03407011>.
- [8] K. Tit, T. Furon, M. Rousset and L.-M. Traonouez. ‘Évaluation statistique efficace de la robustesse de classifieurs’. In: CAID 2021 - Conference on Artificial Intelligence for Defense. Actes de la conférence CAID 2021. Rennes, France, Nov. 2021, pp. 1–11. URL: <https://hal.archives-ouvertes.fr/hal-03462156>.

**Reports & preprints**

- [9] A. Cuillery and F. Le Gland. *Particle filters with auxiliary Markov transition. Application to crossover and to multitarget tracking*. 6th Dec. 2021. URL: <https://hal.inria.fr/hal-03467603>.
- [10] C. Elvira and C. Herzet. *Safe rules for the identification of zeros in the solution of the SLOPE problem*. 25th Oct. 2021. URL: <https://hal.archives-ouvertes.fr/hal-03400322>.
- [11] T. Guyard, C. Herzet and C. Elvira. *Node-screening tests for  $l_0$ -penalized least-squares problem with supplementary material*. 1st Dec. 2021. URL: <https://hal.inria.fr/hal-03462171>.
- [12] T. Guyard, C. Herzet and C. Elvira. *Screen & relax: accelerating the resolution of Elastic-Net by safe identification of the solution support*. 1st Dec. 2021. URL: <https://hal.inria.fr/hal-03462191>.