

The Inria logo is written in a red, cursive script.

IN PARTNERSHIP WITH:  
**CNRS**

**Université Rennes 1**

Activity Report 2019

**Project-Team SIMSMART**

**SIMulating Stochastic Models with pARTicles**

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

RESEARCH CENTER  
**Rennes - Bretagne-Atlantique**

THEME  
**Stochastic approaches**



## Table of contents

<b>1. Team, Visitors, External Collaborators</b> .....	<b>1</b>
<b>2. Overall Objectives</b> .....	<b>2</b>
<b>3. Research Program</b> .....	<b>2</b>
<b>4. Application Domains</b> .....	<b>4</b>
4.1. Domain 1 – Computational Physics	4
4.2. Domain 2 – Meteorology	4
<b>5. New Results</b> .....	<b>5</b>
5.1. Objective 1 – Rare events simulation	5
5.2. Objective 2 – High dimensional filtering	5
5.3. Objective 3 – Non-parametric statistics	6
5.4. Objective 4 – Model Reduction	6
5.5. Miscellaneous	6
<b>6. Bilateral Contracts and Grants with Industry</b> .....	<b>7</b>
6.1. Bilateral Contracts with Industry	7
6.2. Bilateral Grants with Industry	7
<b>7. Partnerships and Cooperations</b> .....	<b>7</b>
7.1. Regional Initiatives	7
7.2. National Initiatives	7
7.3. European Initiatives	8
7.3.1. FP7 & H2020 Projects	8
7.3.2. Collaborations with Major European Organizations	8
7.4. International Initiatives	8
<b>8. Dissemination</b> .....	<b>8</b>
8.1. Promoting Scientific Activities	8
8.2. Teaching - Supervision - Juries	9
8.2.1. Teaching	9
8.2.2. Supervision	9
8.2.3. PhD defended	10
8.2.4. Juries	10
8.3. Popularization	10
<b>9. Bibliography</b> .....	<b>10</b>



# Project-Team SIMSMART

*Creation of the Team: 2018 January 01, updated into Project-Team: 2019 January 01*

## Keywords:

### Computer Science and Digital Science:

- 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
  - A6.3. - 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
    - A6.5.5. - Chemistry

### Other Research Topics and Application Domains:

- B1. - Life sciences
  - B1.1. - Biology
    - B1.1.1. - Structural biology
- B3. - Environment and planet
  - B3.2. - Climate and meteorology
- B4. - Energy
  - B4.2. - Nuclear Energy Production
    - B4.2.1. - Fission

## 1. Team, Visitors, External Collaborators

### Research Scientists

Mathias Rousset [Team leader, Inria, Researcher, HDR]  
Frédéric Céro [Inria, Researcher]  
Cédric Herzet [Inria, Researcher]  
Patrick Héas [Inria, Researcher]  
François Le Gland [Inria, Senior Researcher]

### Faculty Member

Valérie Monbet [Univ Rennes, Professor, HDR]

#### **PhD Students**

Audrey Cuillery [DCNS Group, granted by CIFRE]

Marie Morvan [Univ Rennes]

Gabriel Jouan [Univ Rennes, Scalian, granted by CIFRE]

Esso-Ridah Bleza [Univ Bretagne Sud, Janasense, granted by CIFRE]

Said Obakrim [Univ Rennes, Ifremer]

#### **Intern and Apprentice**

Benjamin Dufee [Inria, from Mar 2019 until Jun 2019]

#### **Administrative Assistant**

Fabienne Cuyollaa [Inria, Administrative Assistant]

#### **External Collaborator**

Arnaud Guyader [Univ Pierre et Marie Curie, HDR]

## **2. Overall Objectives**

### **2.1. 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**

### **3.1. 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's 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:

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

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.

For simplicity, we have classified SIMSMART research into the following already mentioned four main objectives.

1. Objective 1: Rare events and random simulation, which mainly encompass item (i).
2. Objective 2: High dimension and advanced particle filtering, which encompass item (iv).
3. Objective 3: Non-parametric inference, which mainly encompass item (ii) and (vi).
4. Objective 4: Model reduction, which mainly encompasses item (vi).

Uncertainty quantification (item (iii)) in fact underlies each aspect since we are mainly interested in Monte Carlo approaches, so that uncertainty can be *modeled by an initial random variable and be incorporated in the state space of the physical model*.

## 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 MATHERIALS project-teams.

### 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.* <https://www.metoffice.gov.uk/research/modelling-systems/unified-model/>). 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).

## 5. New Results

### 5.1. Objective 1 – Rare events simulation

In [2], we present a short historical perspective of the importance splitting approach to simulate and estimate rare events, with a detailed description of several variants. We then give an account of recent theoretical results on these algorithms, including a central limit theorem for Adaptive Multilevel Splitting (AMS). Considering the asymptotic variance in the latter, the choice of the importance function, called the reaction coordinate in molecular dynamics, is also discussed. Finally, we briefly mention some worthwhile applications of AMS in various domains.

Adaptive Multilevel Splitting (AMS for short) is a generic Monte Carlo method for Markov processes that simulates rare events and estimates associated probabilities. Despite its practical efficiency, there are almost no theoretical results on the convergence of this algorithm. In [1], we prove both consistency and asymptotic normality results in a general setting. This is done by associating to the original Markov process a level-indexed process, also called a stochastic wave, and by showing that AMS can then be seen as a Fleming-Viot type particle system. This being done, we can finally apply general results on Fleming-Viot particle systems that we have recently obtained. In [1] we extend the central limit theorem to the case of synchronized branchings, where re-sampling of particles is performed after any given number of particles have been killed. The result is obtained in the generic case of Fleming-Viot particle systems.

Probability measures supported on submanifolds can be sampled by adding an extra momentum variable to the state of the system, and discretizing the associated Hamiltonian dynamics with some stochastic perturbation in the extra variable. In order to avoid biases in the invariant probability measures sampled by discretizations of these stochastically perturbed Hamiltonian dynamics, a Metropolis rejection procedure can be considered. The so-obtained scheme belongs to the class of generalized Hybrid Monte Carlo (GHMC) algorithms. In [5], we show here how to generalize to GHMC a procedure suggested by Goodman, Holmes-Cerfon and Zappa for Metropolis random walks on submanifolds, where a reverse projection check is performed to enforce the reversibility of the algorithm for large timesteps and hence avoid biases in the invariant measure. We also provide a full mathematical analysis of such procedures, as well as numerical experiments demonstrating the importance of the reverse projection check on simple toy examples.

In [24], we consider Langevin processes, which are widely used in molecular simulation to compute reaction kinetics using rare event algorithms. We prove convergence in distribution in the overdamped asymptotics. The proof relies on the classical perturbed test function (or corrector) method, which is used both to show tightness in path space, and to identify the extracted limit with a martingale problem. The result holds assuming the continuity of the gradient of the potential energy, and a mild control of the initial kinetic energy.

### 5.2. Objective 2 – High dimensional filtering

The work presented in [10] is about solutions for 2D multitarget tracking from image observations including its application on radar data and results on simulated data. Tracking from image observations rather than from detected points is often referred to as track-before-detect (TBD). The objective is to capture targets with a low signal-to-noise ratio (SNR) which would not be detected or tracked after data thresholding. The highly nonlinear filtering equations are approximated using a particle filter implementation. This nonlinearity emerges from the observation function which relies on the radar treatment chain, involving matched filtering of a chirp signal, and beamforming achieved from a linear phased array, leading to the raw data image. Amplitudes and phases of target-returned signals are considered as temporally fluctuating, random and unknown, creating non-deterministic contributions of the targets to the signal to deal with.

### 5.3. Objective 3 – Non-parametric statistics

Production forecast errors are the main hurdle to integrate variable renewable energies into electrical power systems. Regardless of the technique, these errors are inherent in the forecast exercise, although their magnitude significantly vary depending on the method and the horizon. As power systems have to balance out these errors, their dynamic and stochastic modeling is valuable for the real time operation. The study in [23] proposes a Markov Switching Auto Regressive – MS-AR – approach. After having validated its statistical relevance, this model is used to solve the problem of the optimal management of a storage associated with a wind power plant when this virtual power plant must respect a production commitment.

### 5.4. Objective 4 – Model Reduction

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 year, we made several contributions to these subtasks. In most of our contributions, we deviated from the standard working hypothesis involving a linear subspace surrogate.

In a first group of publications, we concentrated our attention on the so-called “sparse” low-dimensional model. In this context, we have proposed several new algorithmic solutions to decrease the computational complexity associated to projection onto this low-dimensional model. These methodologies take place in the context of “screening” procedures for LASSO. We first introduced a new screening strategy, dubbed “joint screening test”, which allows the rejection of a set of atoms by performing one single test, see [4]. Our approach enables to find good compromises between complexity of implementation and effectiveness of screening. Second, we proposed two new methods to decrease the computational cost inherent to the construction of the (so-called) “safe region”. Our numerical experiments show that the proposed procedures lead to significant computational gains as compared to standard methodologies, see [11]. We finally showed in another work that the main concepts underlying screening procedures can be extended to different families of convex optimization problems, see [22].

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 its values in some continuously-valued domain. In this context, we contributed to the theoretical characterization of the performance of some 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 [12], [17], [21]. We also provided a new connection between two popular low-rank approximations of continuous dictionaries, namely the “polar” and “SVD” approximations, see [9].

The tools exploited in the field of model-order reduction and sparsity have found some particular applicative field in geophysics and fluid mechanics. In [8], [7], we derived procedures based on sparse representations to localize the positions of particles in a moving fluid. In [16], [15], [14], [26], [27], we designed learning methodologies to learn the dynamical model underlying a set of observed data.

### 5.5. Miscellaneous

In [25], we devise methods of variance reduction for the Monte Carlo estimation of an expectation of the type  $\mathbb{E}[\phi(X, Y)]$ , when the distribution of  $X$  is exactly known. The key general idea is to give each individual of a sample a weight, so that the resulting weighted empirical distribution has a marginal with respect to the variable  $X$  as close as possible to its target. We prove several theoretical results on the method, identifying settings where the variance reduction is guaranteed. We perform numerical tests comparing the methods and demonstrating their efficiency.

In [6], we consider the problem of predicting a categorical variable based on groups of inputs. Some methods have already been proposed to elaborate classification rules based on groups of variables (e.g. group lasso for logistic regression). However, to our knowledge, no tree-based approach has been proposed to tackle this issue. Here, we propose the Tree Penalized Linear Discriminant Analysis algorithm (TPLDA), a new-tree based approach which constructs a classification rule based on groups of variables.

## 6. Bilateral Contracts and Grants with Industry

### 6.1. Bilateral Contracts with Industry

1. **Scalian Alyotech**, through the CIFRE PhD project of Gabriel Jouan, dedicated to weather forecast corrections.
2. **Naval Group Research**, through the CIFRE PhD project of Audrey Cuillery dedicated to Bayesian tracking.
3. **Eau du Ponant**, through the R&D project MEDISA (<https://www.eauduponant.fr/fr/actualite/lancement-du-projet-de-rd-medisa>) on water industry.
4. **Cooper Standard**, Machine Learning for joints design.

### 6.2. Bilateral Grants with Industry

1. **EURAMED** (a Euro-Mediterranean Cooperation Initiative, which aims to develop an Internet-based, multi-parametric electronic platform for optimum design of desalination plants, supplied by Renewable Energy Sources (RES). PI: E. Koutroulis (GREECE).

## 7. Partnerships and Cooperations

### 7.1. Regional Initiatives

**Inter-Labex SEACS:** V. Monbet, F. Le Gland, C. Herzet and Thi Tuyet Trang Chau (PhD student) are part of the *inter Labex Cominlabs-Lebesgue-Mer SEACS*, <http://www.seacs.cominlabs.ueb.eu/fr>, which stands for Stochastic model-dAta-Coupled representationS for the analysis, simulation and reconstruction of upper ocean dynamics. This project which concerns mainly Objectives 2 and 3, aims at exploring novel statistical and stochastic methods to address the emulation, reconstruction and forecast of fine-scale upper ocean dynamics.maths-computer-sea science for ocean dynamics.

**CMEMS 3DA (2018-2019):** C. Herzet is part of the project *CMEMS 3DA* on data assimilation of oceanographic events with non-parametric data assimilation methods. The goal of the project is to demonstrate the relevance of data-driven strategies to improve satellite derived interpolated products and especially the geostrophic surface currents. The project is made in collaboration with IMT Atlantique Brest, Ifremer and the Institut of Geosciences and Environment in Grenoble.

**Action Exploratoire – Labex Cominlabs:** C. Herzet is part of a project on sparse representations in continuous dictionaries. Partners: R. Gribonval (Inria Rennes PANAMA), A. Drémeau (IMT Atlantique) and P. Tando (IMT Atlantique).

### 7.2. National Initiatives

#### 7.2.1. ANR

**ANR BECOSE (2016-2020):** Beyond Compressive Sensing: Sparse approximation algorithms for ill-conditioned inverse problems.

Cédric Herzet is part of the BECOSE project. The BECOSE project aims to extend the scope of sparsity techniques much beyond the academic setting of random and well-conditioned dictionaries. In particular, one goal of the project is to step back from the popular L1-convexification of the sparse representation problem and consider more involved nonconvex formulations, both from a methodological and theoretical point of view. The algorithms will be assessed in the context of tomographic Particle Image Velocimetry (PIV), a rapidly growing imaging technique in fluid mechanics that will have strong impact in several industrial sectors including environment, automotive and aeronautical industries.

**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).

## 7.3. European Initiatives

### 7.3.1. FP7 & H2020 Projects

**ERC MsMaths (2015-2019):** M. Rousset is part of *ERC MsMaths* on molecular simulation (PI T. Lelièvre). With the development of large-scale computing facilities, simulations of materials at the molecular scale are now performed on a daily basis. The objective of the MSMATH ERC project is to develop and study efficient algorithms to simulate such high-dimensional systems over very long, macroscopic times. ERC MsMaths especially focus on the computational issues related to 'metastable' states, that is to say specific molecular configurations that do evolve only on very large time scales. This results in a multi-timescale computational bottleneck that needs to be addressed by specific algorithms.

### 7.3.2. Collaborations with Major European Organizations

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 instrument (IRS on MTG-S, developed by ESA and IASI-NG on Metop-SG developed by CNES).

## 7.4. International Initiatives

### 7.4.1. Participation in Other International Programs

**ECOS ARGENTINE (2018-2021):** V. Monbet has obtained a funding program through the ECOS Sud - MINCYT initiative (<http://www.univ-paris13.fr/cofecub-ecos/>). The program involves a collaboration with the French-Argentinian Climate Institute (<http://www.cima.fcen.uba.ar/UMI/>), and focuses on non-parametric, analog methods, combined with data assimilation techniques to reconstruct complex meteorological dynamics (Objective 3).

## 8. Dissemination

### 8.1. Promoting Scientific Activities

#### 8.1.1. Member of the Organizing Committees

Cédric Herzet is part of the organizing committee of the iTwist'20 Workshop.

## 8.2. Teaching - Supervision - Juries

### 8.2.1. Teaching

Cédric Herzet has given:

- INSA RENNES, 5<sup>ème</sup> année de l'option Génie Mathématique, cours de Parcimonie en traitement du signal et des images, 10h de cours magistraux + responsable du module
- Ensai RENNES, Master international « Smart Data », cours « Foundations of Smart Sensing », 9h de cours magistraux
- Ensai RENNES, Master international « Smart Data », cours « Advanced topics in Smart Sensing », 3h de cours magistraux
- Ensai RENNES, Master 1, « Régression pénalisée et sélection de modèles », cours « Advanced topics in Smart Sensing », 6h de cours magistraux + 6 TPs + responsable du module

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 GM/AROM (risk analysis, optimization and modeling) major in mathematical engineering,
- 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,
- and a course on **Kalman filtering and hidden Markov models**, at université de Rennes 1, within the SISEA (signal, image, systèmes embarqués, automatique, école doctorale MATISSE) track of the master in electronical engineering and telecommunicationst.

### 8.2.2. Supervision

Cédric Herzet has supervised:

- Soufiane Ait Tilat, PhD, co-supervision with Frédéric Champagnat (Onera, Palaiseau)
- Milan Courcoux-Caro, PhD, co-supervision with Charles Vanwynsberghe (ENSTA Bretagne) and Alexandre Baussard (IUT de Troyes),
- Clément Elvira, postdoc, co-supervision with Rémi Gribonval (Inria Rennes) and Charles Soussen (CentraleSupélec)
- Clément Dorffer, postdoc, co-supervision with Angélique Drémeau (ENSTA Rennes)

Mathias Rousset has supervised:

- Benjamin Dufée, master 2 (with Frédéric Cérou).

François Le Gland has supervised

- Audrey Cuillery, PhD, provisional title: *Bayesian tracking from raw data*, université de Rennes 1, started in April 2016, expected defense in early 2020, funding: CIFRE grant with Naval Group, co-direction: Dann Laneuville (Naval Group, Nantes).

V. Monbet has supervised

- Gabriel Jouan, PhD, Univ Rennes, Scalian, granted by CIFRE.
- Esso-Ridah Bleza, PhD, Univ Bretagne Sud, Janasense, granted by CIFRE.
- Said Obakrim, PhD, Univ Rennes, Ifremer.

### 8.2.3. PhD defended

- Y. Xu (sup.: M. Rousset and P.A. Zitt): Weak over-damped asymptotic and variance reduction
- T.T.T. Chau (sup: V. Monbet): Non-parametric methodologies for reconstruction and estimation in nonlinear state-space models
- M. Morvan (sup: V. Monbet): Modèles de régression pour données fonctionnelles hétérogènes. Application à la modélisation de données de spectrométrie dans le moyen infrarouge

### 8.2.4. Juries

V. Monbet has been a member of the following juries:

- Anders Hildeman - On flexible random field models for spatial statistics: Spatial mixture models and deformed SPDE models, Chalmers University, Sweden.
- Shuaitao Wang - Simulation du métabolisme de la Seine par assimilation de données en continu" , Mines Paris-Tech (reviewer)
- Alban FARCHI - Localisation des méthodes d'assimilation de données d'ensemble, ENPC

François Le Gland has been a member of the following juries:

- (Reviewer) Julien Lesouple (université de Toulouse, adviser : Jean-Yves Tournet)
- Émilien Flayac (université Paris–Saclay, Orsay, advisers : Frédéric Jean and Karim Dahia)
- Thi Tuyet Trang Chau (université de Rennes 1, advisers : Valérie Monbet and Pierre Ailliot).

## 8.3. Popularization

### 8.3.1. Interventions

Patrick Héas has made a CS workshop for middle school students based on the ideas of 'Computer Science Unplugged'.

## 9. Bibliography

### Publications of the year

#### Articles in International Peer-Reviewed Journals

- [1] F. CÉROU, B. DELYON, A. GUYADER, M. ROUSSET. *On the Asymptotic Normality of Adaptive Multilevel Splitting*, in "SIAM/ASA Journal on Uncertainty Quantification", 2019, vol. 7, n<sup>o</sup> 1, pp. 1-30, <https://arxiv.org/abs/1804.08494> - 38 pages, 5 figures [DOI : 10.1137/18M1187477], <https://hal.archives-ouvertes.fr/hal-01774856>
- [2] F. CÉROU, A. GUYADER, M. ROUSSET. *Adaptive Multilevel Splitting: Historical Perspective and Recent Results*, in "Chaos", April 2019, vol. 29, n<sup>o</sup> 4, pp. 1-32 [DOI : 10.1063/1.5082247], <https://hal.inria.fr/hal-01936611>
- [3] C. HERZET, M. M. DIALLO. *Performance guarantees for a variational "multi-space" decoder*, in "Advances in Computational Mathematics", 2019, pp. 1-23, forthcoming, <https://hal.inria.fr/hal-02394378>
- [4] C. HERZET, C. DORFFER, A. DRÉMEAU. *Gather and Conquer: Region-Based Strategies to Accelerate Safe Screening Tests*, in "IEEE Transactions on Signal Processing", June 2019, vol. 67, n<sup>o</sup> 12, pp. 3300-3315 [DOI : 10.1109/TSP.2019.2914885], <https://hal-ensta-bretagne.archives-ouvertes.fr/hal-02303198>

- [5] T. LELIÈVRE, M. ROUSSET, G. STOLTZ. *Hybrid Monte Carlo methods for sampling probability measures on submanifolds*, in "Numerische Mathematik", June 2019, vol. 143, n<sup>o</sup> 2, pp. 379–421, <https://arxiv.org/abs/1807.02356> [DOI : 10.1007/s00211-019-01056-4], <https://hal.archives-ouvertes.fr/hal-01832820>
- [6] A. POTERIE, J.-F. DUPUY, V. MONBET, L. ROUVIÈRE. *Classification tree algorithm for grouped variables*, in "Computational Statistics", 2019, pp. 1613–1648, forthcoming [DOI : 10.1007/s00180-019-00894-y], <https://hal.archives-ouvertes.fr/hal-01623570>

### International Conferences with Proceedings

- [7] S. AIT TILAT, F. CHAMPAGNAT, C. HERZET. *A new sparsity based particle image reconstruction approach for particle detection*, in "ISPIV 2019 - 13th International Symposium on Particle Image Velocimetry", Munich, Germany, July 2019, pp. 1-10, <https://hal.inria.fr/hal-02389648>
- [8] S. AIT TILAT, C. HERZET, F. CHAMPAGNAT. *Détection et localisation en imagerie de particules par approximation de Taylor de dictionnaires continus*, in "GRETSI 2019", Lille, France, August 2019, pp. 1-4, <https://hal.archives-ouvertes.fr/hal-02362086>
- [9] F. CHAMPAGNAT, C. HERZET. *Atom selection in continuous dictionaries : reconciling polar and SVD approximations*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 5516-5520 [DOI : 10.1109/ICASSP.2019.8683402], <https://hal.inria.fr/hal-02059703>
- [10] A. CUILLERY, F. LE GLAND. *Track-before-detect on radar image observation with an adaptive auxiliary particle filter*, in "FUSION 2019 - 22nd International Conference on Information Fusion", Ottawa, Canada, July 2019, <https://hal.inria.fr/hal-02413324>
- [11] C. DORFFER, C. HERZET, A. DRÉMEAU. *Region-based relaxations to accelerate greedy approaches*, in "27th European Signal Processing Conference, EUSIPCO 2019", La Corogne, Spain, September 2019, <https://hal.inria.fr/hal-02059649>
- [12] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN. *Uniform k-step recovery with CMF dictionaries*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations", Toulouse, France, July 2019, pp. 1-2, <https://hal.inria.fr/hal-02157561>
- [13] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *OMP and continuous dictionaries: Is k-step recovery possible ?*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 1-5 [DOI : 10.1109/ICASSP.2019.8683617], <https://hal.archives-ouvertes.fr/hal-02049486>
- [14] S. OUALA, R. FABLET, C. HERZET, B. CHAPRON, A. PASCUAL, F. COLLARD, L. GAULTIER. *Learning stochastic representations of geophysical dynamics*, in "44th International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2019", Brighton, United Kingdom, May 2019, <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02005403>
- [15] S. OUALA, R. FABLET, C. HERZET, L. DRUMETZ, B. CHAPRON, A. PASCUAL, F. COLLARD, L. GAULTIER. *Sea surface dynamics reconstruction using neural networks based kalman filter*, in "IGARSS 2019 - International Geoscience and remote Sensing Symposium", Yokohama, Japan, July 2019, pp. 1-5 [DOI : 10.13140/RG.2.2.11461.09446], <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02285697>

- [16] S. OUALA, D. NGUYEN, L. DRUMETZ, C. HERZET, A. PASCUAL, L. GAULTIER, F. COLLARD, B. CHAPRON, R. FABLET. *Learning ocean dynamical priors from noisy data using assimilation-derived neural nets*, in "IGARSS 2019 - International Geoscience and remote Sensing Symposium", Yokohama, Japan, July 2019, pp. 1-3 [DOI : 10.13140/RG.2.2.35788.05767], <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02285693>

### National Conferences with Proceedings

- [17] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *Identification de supports en  $k$  étapes avec OMP pour les dictionnaires continus*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, pp. 1-4, <https://hal.inria.fr/hal-02157571>

### Conferences without Proceedings

- [18] G. JOUAN, A. CUZOL, V. MONBET, G. MONNIER. *Weather types prediction at medium-range from ensemble forecasts*, in "9th International workshop on Climate Informatics", Paris, France, October 2019, <https://hal.archives-ouvertes.fr/hal-02425230>

### Other Publications

- [19] P. AILLIOT, M. BOUTIGNY, E. KOUTROULIS, A. MALISOVAS, V. MONBET. *Stochastic Weather Generator for the Design and Reliability Evaluation of Desalination Systems with Renewable Energy Sources*, December 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02425916>
- [20] F. CÉROU, A. GUYADER, M. ROUSSET. *On Synchronized Fleming-Viot Particle Systems*, November 2019, <https://arxiv.org/abs/1911.05366> - 39 pages, 3 figures, <https://hal.archives-ouvertes.fr/hal-02362847>
- [21] C. ELVIRA, R. GRIBONVAL, C. SOUSSEN, C. HERZET. *When does OMP achieve support recovery with continuous dictionaries?*, April 2019, working paper or preprint, <https://hal.inria.fr/hal-02099464>
- [22] C. ELVIRA, C. HERZET. *Safe Squeezing for Antispase Coding*, November 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02368134>
- [23] R. LE GOFF LATIMIER, E. LE BOUËDEC, V. MONBET. *Modeling of the wind power forecast errors and associated optimal storage strategy*, June 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02151567>
- [24] M. ROUSSET, Y. XU, P.-A. ZITT. *A Weak Overdamped Limit Theorem for Langevin Processes*, March 2019, <https://arxiv.org/abs/1709.09866> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01596954>
- [25] M. ROUSSET, Y. XU, P.-A. ZITT. *Reducing Variance by Reweighting Samples*, March 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01925646>
- [26] Y. ZHEN, P. TANDEO, S. LEROUX, J. LE SOMMER, P. AILLIOT, R. FABLET, B. CHAPRON, C. HERZET, T. PENDUFF, J. VERRON, S. METREF. *Analog forecasting (AnDA) and large-ensemble ocean simulations to improve satellite-derived gridded products*, May 2019, 1 p. , OceanPredict '19 - GODAE OceanView Symposium, Poster, <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02156704>
- [27] Y. ZHEN, P. TANDEO, S. LEROUX, J. LE SOMMER, S. METREF, P. AILLIOT, R. FABLET, B. CHAPRON, C. HERZET, T. PENDUFF, J. VERRON. *3DA: Data-Driven Data Assimilation Applications to the Lorenz*



*system and simulated sea surface heights*, April 2019, 1 p. , EGU 2019 : European Geosciences Union, Poster,  
<https://hal-imt-atlantique.archives-ouvertes.fr/hal-02108347>