Activity Report 2018

Team SIMSMART

SIMulating Stochastic Models with pARTicles

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Stochastic approaches

Table of contents

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 Cérou [Inria, Researcher]
Cédric Herzet [Inria, Researcher, from May 2018]
Patrick Héas [Inria, Researcher]
François Le Gland [Inria, Senior Researcher]

Faculty Member

Valérie Monbet [Univ de Rennes I, Professor, HDR]

PhD Student

Audrey Cuillery [Naval Group]

Administrative Assistant

Fabienne Cuyollaa [Inria]

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 phenomenons, 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. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

The contribution [10] received the "best poster award" of the conference Curves and Surfaces 2018.

6. New Results

6.1. Objective 1 – Rare events simulation

In [16], we present a short historical perpective 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 [15], 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.

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 [21], 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.

Feynman-Kac semigroups appear in various areas of mathematics: non-linear filtering, large deviations theory, spectral analysis of Schrodinger operators among others. Their long time behavior provides important information, for example in terms of ground state energy of Schrodinger operators, or scaled cumulant generating function in large deviations theory. In [17], we propose a simple and natural extension of the stability of Markov chains for these non-linear evolutions. As other classical ergodicity results, it relies on two assumptions: a Lyapunov condition that induces some compactness, and a minorization condition ensuring some mixing. We show that these conditions are satisfied in a variety of situations. We also show that our technique provides uniform in the time step convergence estimates for discretizations of stochastic differential equations.

6.2. Objective 2 – High dimensional and advanced particle filtering

Existing filtering based structural health monitoring (SHM) algorithms assume constant noise environment which does not always conform to the reality as noise is hardly stationary. Thus to ensure optimal solution even with non-stationary noise processes, the assumed statistical noise models have to be updated periodically. [8] incorporates a modification in the existing Interacting Particle-Kalman Filter (IPKF) to enhance its detection capability in presence of non-stationary noise processes. To achieve noise adaptability, the proposed algorithm recursively estimates and updates the current noise statistics using the post-IPKF residual uncertainty in prediction as a measurement which in turn enhances the optimality in the solution as well. Further, this algorithm also attempts to mitigate the ill effects of abrupt change in noise statistics which most often deteriorates/ diverges the estimation. For this, the Kalman filters (KF) within the IPKF have been replaced with a maximum Correntropy criterion (MCC) based KF that, unlike regular KF, takes moments beyond second order into consideration. A Gaussian kernel for MCC criterion is employed to define a correntropy index that controls the update in state and noise estimates in each recursive steps. Numerical experiments on an eight degrees-of-freedom system establish the potential of this algorithm in real field applications.

Standard filtering techniques for structural parameter estimation assume that the input force is either known or can be replicated using a known white Gaussian model. Unfortunately for structures subjected to seismic excitation, the input time history is unknown and also no previously known representative model is available. This invalidates the aforementioned idealization. To identify seismic induced damage in such structures using filtering techniques, force must therefore also be estimated. In [5], the input force is considered to be an additional state that is estimated in parallel to the structural parameters. Two concurrent filters are employed for parameters and force respectively. For the parameters, an interacting Particle-Kalman filter is used to target systems with correlated noise. Alongside this, a second filter is used to estimate the seismic force acting on the structure. In the proposed algorithm, the parameters and the inputs are estimated as being conditional on each other, thus ensuring stability in the estimation. The proposed algorithm is numerically validated on a sixteen degrees-of-freedom mass-spring-damper system and a five-story building structure. The stability of the proposed filter is also tested by subjecting it to a sufficiently long measurement time history. The estimation results confirm the applicability of the proposed algorithm.

6.3. Objective 3 – Non-parametric inference

The forecasting and reconstruction of ocean and atmosphere dynamics from satellite observation time series are key challenges. While model-driven representations remain the classic approaches, data-driven representations become more and more appealing to benefit from available large-scale observation and simulation datasets. In [12], [13] and [4], we investigate the relevance of recently introduced neural network representations for the forecasting and assimilation of geophysical fields from satellite-derived remote sensing data. As a case-study, we consider satellite-derived Sea Surface Temperature time series off South Africa, which involves intense and complex upper ocean dynamics. Our numerical experiments report significant improvements in terms of reconstruction performance compared with operational and state-of-the-art schemes.

Data assimilation methods aim at estimating the state of a system by combining observations with a physical model. When sequential data assimilation is considered, the joint distribution of the latent state and the observations is described mathematically using a state-space model, and filtering or smoothing algorithms are used to approximate the conditional distribution of the state given the observations. The most popular algorithms in the data assimilation community are based on the Ensemble Kalman Filter and Smoother (EnKF/EnKS) and its extensions. In [14], we investigate an alternative approach where a Conditional Particle Filter (CPF) is combined with Backward Simulation (BS). This allows to explore efficiently the latent space and simulate quickly relevant trajectories of the state conditionally to the observations. We also tackle the difficult problem of parameter estimation. Indeed, the models generally involve statistical parameters in the physical models and/or in the stochastic models for the errors. These parameters strongly impact the results of the data assimilation algorithm and there is a need for an efficient method to estimate them. Expectation-Maximization (EM) is the most classical algorithm in the statistical literature to estimate the parameters in models with latent variables. It consists in updating sequentially the parameters by maximizing a likelihood function where the state is approximated using a smoothing algorithm. In this paper, we propose an original Stochastic Expectation-Maximization (SEM) algorithm combined to the CPF-BS smoother to estimate the statistical parameters. We show on several toy models that this algorithm provides, with reasonable computational cost, accurate estimations of the statistical parameters and the state in highly nonlinear state-space models, where the application of EM algorithms using EnKS is limited. We also provide a Python source code of the algorithm.

6.4. Objective 4 – Model reduction

In [19], [7], we propose new methodologies to decrease the computational cost of safe screening tests for LASSO. We first introduce a new screening strategy, dubbed "joint screening test", which allows the rejection of a set of atoms by performing one single test. Our approach enables to find good compromises between complexity of implementation and effectiveness of screening. Second, we propose 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.

Model-order reduction methods tackle the following general approximation problem: find an "easily-computable" but accurate approximation of some target solution h. In order to achieve this goal, standard methodologies combine two main ingredients: i) a set of problem-specific constraints; ii) some "simple" prior model on the set of target solutions. The most common prior model encountered in the literature assume that the target solution h is "close" to some low-dimensional subspace. Recently, several contributions have shown that refined prior models (based on a set of embedded approximation subspaces) may lead to enhanced approximation performance. Unfortunately, to date, no theoretical results have been derived to support the good empirical performance observed in these contributions. The goal of [18] is to fill this gap. More specifically, we provide a mathematical characterization of the approximation performance achievable by some particular "multi-space" decoder and emphasize that, in some specific setups, this "multi-space" decoder has provably better recovery guarantees than its standard counterpart based on a single approximation subspace.

In [20], we deal with the estimation of rare event probabilities using importance sampling (IS), where an *optimal* proposal distribution is computed with the cross-entropy (CE) method. Although, IS optimized with the CE method leads to an efficient reduction of the estimator variance, this approach remains unaffordable for problems where the repeated evaluation of the score function represents a too intensive computational effort. This is often the case for score functions related to the solution of parametric partial differential equations (PPDE) with random inputs. This work proposes to alleviate computation by adapting a score function approximation along the CE optimization process. The score function approximation is obtained by selecting the surrogate of lowest dimensionality, whose accuracy guarantees to pass the current CE optimization stage. The adaptation of the surrogate relies on certified upper bounds on the error norm. An asymptotic analysis provides some theoretical guarantees on the efficiency and convergence of the proposed algorithm. Numerical results demonstrate the gain brought by the adaptive method in the context of pollution alerts and a system modelled by a PPDE.

In [2], we deal with model order reduction of PPDE. We consider the specific setup where the solutions of the PPDE are only observed through a partial observation operator and address the task of finding a good approximation subspace of the solution manifold. We provide and study several tools to tackle this problem. We first identify the best worst-case performance achievable in this setup and propose simple procedures to approximate this optimal solution. We then provide, in a simplified setup, a theoretical analysis relating the achievable reduction performance to the choice of the observation operator and the prior knowledge available on the solution manifold.

In [3], we deal with model order reduction of parametrical dynamical systems. We consider the specific setup where the distribution of the system's trajectories is unknown but the following two sources of information are available: *(i)* some "rough" prior knowledge on the system's realisations; *(ii)* a set of "incomplete" observations of the system's trajectories. We propose a Bayesian methodological framework to build reduced-order models (ROMs) by exploiting these two sources of information. We emphasise that complementing the prior knowledge with the collected data provably enhances the knowledge of the distribution of the system's trajectories. We then propose an implementation of the proposed methodology based on Monte-Carlo methods. In this context, we show that standard ROM learning techniques, such e.g. Proper Orthogonal Decomposition or Dynamic Mode Decomposition, can be revisited and recast within the probabilistic framework considered in this paper. We illustrate the performance of the proposed approach by numerical results obtained for a standard geophysical model.

6.5. Miscellaneous

In [22], 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.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts or Grants with Industry (Private Sector)

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.

7.2. Bilateral Contracts or Grants with Industry (Public Sector)

1. **CEA LETI** on indoor navigation (particle filtering) through the CEA PhD grant of Kersane Zoubert–Ousseni.
2. **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).

8. Partnerships and Cooperations

8.1. 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, https://seacs.cominlabs.u-bretagneloire.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 Institue 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. Tandeo (IMT Atlantique).

8.2. National Initiatives

8.2.1. ANR

ANR COSMOS (2014-2018): F. Cérou and A. Guyader are part of *ANR Cosmos* on molecular simulation and statistics (PI G. Stoltz of MATHERIALS). COSMOS aims at developing numerical techniques dedicated to the sampling of high-dimensional probability measures describing a system of interest. There are two application fields of interest: computational statistical physics (a field also known as molecular simulation), and computational statistics, both sharing a common history and mathematical tools. Our specific role in the project is to study the theoretical aspects of the simulation of reactive trajectories (short trajectories linking two metastable states of the system), which can be viewed as a special type of rare event. These algorithms are then incorporated in molecular simulation softwares by members of MATHERIALS team. They also contribute to popularize them within the wider computational statistical physics community.

ANR Geronimo (2014-2018): C. Herzet (PI) and P. Héas are part of the ANR Geronimo. Its objective is the conception of new techniques for the construction of geophysical reduced-order models (ROMs) from image data. The project both arises from the crucial need of accurate low-order descriptions of highly-complex geophysical phenomena and the recent numerical revolution which has supplied the geophysical scientists with an unprecedented volume of image data. As the output of the project, we devised several methodologies to build effective reduced-order models from data. The research objectives of Objective 4 are in the continuation of the Geronimo project. In particular, we intend to tackle model-order reduction in a probabilistic setup and provide new tools to address the non-linear case.

8.3. European Initiatives

8.3.1. 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.

8.3.2. 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). The work consists in the design of an efficient and physically-based methodology for the estimation of vertically resolved 3D atmospheric motion vector (AMV) fields at various altitude levels. The estimation is based on image sequence observations, depicting temperature, specific humidity or ozone fields at different pressure levels. The final objective is the operational production of 3D AMV fields, which should be used by the different international institutes of meteorology, such as *Meteo France* or the *Met Office*. It is expected that these image-based wind estimates will significantly impact data-assimilation for weather forecasts or climate studies. We mention that the problem of 3D AMVs estimation has several specificities which makes it particularly challenging (high-dimensionality, non-convex and non-differentiable problem). We are working on an overall algorithmic solution to address this problem. Type of collaboration: supervision of an engineer with one or two week-long visits per year. A first prototype (free software licence LGPL 2.1) is currently under evaluation.

8.4. International Initiatives

8.4.1. Participation in Other International Programs

ECOS ARGENTINE (2018-2021): V. Monbet has obtained a funding program through the ECOS Sud - MINCyT intiative (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).

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

- F. Cérou and M. Rousset have been the main organizer and scientific manager of the Workshop Simulation and probability: recent trends (5-8 June, Rennes) within the Thematic Semester 2018 of Labex Lebesgue.
- V.Monbet has organized a Workshop in Rennes "Fonctional Data Analysis 2018", Oct. 24-26, 2018, https://perso.univ-rennes1.fr/valerie.monbet/FDA2018/PEPS_FDA2018_Workshop.html.
- V.Monbet has organized a Workshop in Rennes on Statistics in Meteorology, Nov. 28-30 2018, https://spatiotempmeteo.sciencesconf.org/.

9.1.1.2. Member of the Organizing Committees

Cédric Herzet was part of the Conference Program Committee of the iTwist'18 Workshop

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

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,
- 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 telecommunications,
- and a 3rd year course on hidden Markov models, at Télécom Bretagne, aka IMT Atlantique, Brest.

Cédric Herzet has given

- a Master course at INSA Rennes, option Génie Mathématique, cours de Parcimonie en traitement du signal et des images.
- a Master course at Ensai Rennes, Master international “Smart Data” , cours “Foundations of Smart Sensing”.

9.2.2. PhDs of the team members

- Defended PhD: Kersane Zoubert–Ousseni, CEA LETI, Grenoble. Subject: Off–line indoor navigation [1], supported by a CEA grant, with Christophe Villien (CEA LETI, Grenoble) as co-advisor. From Dec 2014 to Apr 2018. This PhD is related to our applicative/transfer activity on filtering.
- In progress: Thi Tuyet Trang Chau, IRMAR Rennes 1 University. Subject: new statistical methods for missing-data imputation and non-parametric state-space modeling, supported by Labex Lebesgue and Brittany Council, with Pierre Ailliot (UBO) as co-advisor. Since Oct 2015. This PhD is related to Objective 3.
- In progress: Audrey Cuillery, Inria rennes. Subject: Bayesian tracking from raw data, supported by a CIFRE grant from Naval Group (DCNS) Research, with Dann Laneuville (Naval Group) as co-advisor. Since May 2016. This PhD is related to our applicative/transfer activity on filtering.
- In progress: Audrey Poterie, IRMAR and INSA Rennes. Subject: extension of learning methods to grouped variables, supported by a French Research Ministry grant, with Jean-François Dupuy (INSA) and Laurent Rouvière (Rennes 2) as co-advisors. Since Oct 2015. This PhD is related to Objective 3.
- In progress: Yushun Xu, LAMA, Paris-Est University. Subject: Variance reduction and Langevin processes, supported by a French Research Ministry grant, with Pierre-André Zitt (LAMA) as a co-advisor. Since Sept 2015. This PhD is related to Objective 1 and Application 1.
- Starting: Gabriel Jouan, IRMAR, Rennes. Subject: parametric and non-parametric weather forecast corrections from catalogs of weather data records, supported by a CIFRE grant of Scalian Alyotech, with Soulvan Monnier (Scalian Alyothech) as co–advisor. Starting 2018. This PhD is related to our applicative/transfer activity, as well as to Objective 2 and Objective 3.

C. Herzet is the co-supervisor of two PhDs, one with Frédéric Champagnat (Onera, Palaiseau), and one with Charles Vanwynsberghe (ENSTA Bretagne) and Alexandre Baussard (IUT de Troyes).

9.2.3. Juries

- François Le Gland has been a member of the PhD defense committee of Nicolas Merlinge (université Paris Sud and Coventry University, advisers : Hélène Piet–Lahanier and James Brusey).
- Cédric Herzet has been a member of the PhD defense committee of Quentin Denoyelle, « Theoretical and Numerical Analysis of Super-Resolution without Grid », Paris-Dauphine, Thesis director : Gabriel Peyré.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] K. ZOUBERT-OUSSENI. *Post-processing algorithms for indoor localization*, Université Rennes 1, April 2018, https://tel.archives-ouvertes.fr/tel-01838480

Articles in International Peer-Reviewed Journals

[2] C. HERZET, A. DRÉMEAU, P. HÉAS. *Model Reduction from Partial Observations*, in "International Journal for Numerical Methods in Engineering", January 2018, vol. 113, n^{o} 3, pp. 479-511 [*DOI :* 10.1002/NME.5623], https://hal.archives-ouvertes.fr/hal-01394059

[3] P. HÉAS, C. HERZET. *Reduced Modeling of Unknown Trajectories*, in "Archives of Computational Methods in Engineering", January 2018, vol. 25, n^{o} 1, pp. 87-101, https://arxiv.org/abs/1702.08846 [*DOI :* 10.1007/S11831-017-9229-0], https://hal.archives-ouvertes.fr/hal-01490572

[4] S. OUALA, R. FABLET, C. HERZET, B. CHAPRON, A. PASCUAL, F. COLLARD, L. GAULTIER. *Neural-Network-based Kalman Filters for the Spatio-Temporal Interpolation of Satellite-derived Sea Surface Temperature*, in "Remote Sensing", November 2018, https://hal-imt-atlantique.archives-ouvertes.fr/hal-01896654

[5] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. *Seismic-induced damage detection through parallel force and parameter estimation using an improved interacting Particle-Kalman filter*, in "Mechanical Systems and Signal Processing", September 2018, vol. 110, pp. 231 - 247 [*DOI :* 10.1016/J.YMSSP.2018.03.016], https://hal.inria.fr/hal-01806896

International Conferences with Proceedings

[6] R. FABLET, S. OUALA, C. HERZET. *Bilinear residual Neural Network for the identification and forecasting of dynamical systems*, in "EUSIPCO 2018, European Signal Processing Conference", Rome, Italy, September 2018, https://arxiv.org/abs/1712.07003 , https://hal.archives-ouvertes.fr/hal-01686766

[7] C. HERZET, A. DRÉMEAU. *Joint Screening Tests for LASSO*, in "ICASSP 2018 - IEEE International Conference on Acoustic, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 4084-4088 [*DOI :* 10.1109/ICASSP.2018.8462530], https://hal.inria.fr/hal-01739477

[8] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. *Correntropy based IPKF filter for parameter estimation in presence of non-stationary noise process*, in "SAFEPROCESS 2018 - 10th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes", Varsovie, Poland, August 2018, https://hal.inria.fr/hal-01887557

Conferences without Proceedings

[9] C. DORFFER, A. DRÉMEAU, C. HERZET. *Efficient atom selection strategy for iterative sparse approximations*, in "iTWIST 2018 - International Traveling Workshop on Interactions between low-complexity data models and Sensing Techniques", Marseille, France, November 2018, pp. 1-3, https://hal.inria.fr/hal-01937501

[10] C. ELVIRA, R. GRIBONVAL, C. HERZET, C. SOUSSEN. *A case of exact recovery with OMP using continuous dictionaries*, in "CS 2018 - 9th International Conference on Curves and Surfaces", Arcachon, France, June 2018, https://hal.inria.fr/hal-01937532

[11] C. HERZET, M. DIALLO, P. HÉAS. *Beyond Petrov-Galerkin projection by using “multi-space” priors*, in "MoRePas 2018 - Model Reduction for Parametrized Systems IV", Nantes, France, April 2018, https://hal.inria.fr/hal-01937876

[12] S. OUALA, C. HERZET, R. FABLET. *Réseaux de neurones artificiels pour la prédiction et la reconstruction de dynamiques océanographiques*, in "CFPT 2018 - Conférence Française de Photogrammétrie et de Télédétection", Marne-la-Vallée, France, June 2018, https://hal-imt-atlantique.archives-ouvertes.fr/hal-01883211

[13] S. OUALA, C. HERZET, R. FABLET. *Sea surface temperature prediction and reconstruction using patch-level neural network representations*, in "IGARSS 2018 - International Geoscience and Remote Sensing Symposium", Valence, Spain, July 2018, pp. 1-4, https://arxiv.org/abs/1806.00144 , https://hal-imt-atlantique.archives-ouvertes.fr/hal-01883209

Other Publications

[14] T. T. T. CHAU, P. AILLIOT, V. MONBET, P. TANDEO. *An efficient particle-based method for maximum likelihood estimation in nonlinear state-space models*, February 2018, https://arxiv.org/abs/1804.07483 - working paper or preprint, https://hal.inria.fr/hal-01708631

[15] F. CÉROU, B. DELYON, A. GUYADER, M. ROUSSET. *On the Asymptotic Normality of Adaptive Multilevel Splitting*, April 2018, https://arxiv.org/abs/1804.08494 - 38 pages, 5 figures, https://hal.archives-ouvertes.fr/hal-01774856

[16] F. CÉROU, A. GUYADER, M. ROUSSET. *Adaptive Multilevel Splitting: Historical Perspective and Recent Results*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01936611

[17] G. FERRE, M. ROUSSET, G. STOLTZ. *More on the long time stability of Feynman-Kac semigroups*, November 2018, https://arxiv.org/abs/1807.00390 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01911620

[18] C. HERZET, M. DIALLO, P. HÉAS. *An instance optimality property for approximation problems with multiple approximation subspaces*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01913339

[19] C. HERZET, C. DORFFER, A. DRÉMEAU. *Gather and Conquer: Region-based Strategies to Accelerate Safe Screening Tests*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01913331

[20] P. HÉAS. *Adapting Reduced Models in the Cross-Entropy Method*, November 2018, https://arxiv.org/abs/1806.01916 - 25 pages, https://hal.archives-ouvertes.fr/hal-01939804

[21] T. LELIEVRE, M. ROUSSET, G. STOLTZ. *Hybrid Monte Carlo methods for sampling probability measures on submanifolds*, July 2018, https://arxiv.org/abs/1807.02356 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01832820

[22] M. ROUSSET, Y. XU, P.-A. ZITT. *Reducing Variance by Reweighting Samples*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01925646