RESEARCH CENTRE

Inria Center
at Rennes University

IN PARTNERSHIP WITH:
Université Rennes 1, CNRS

2022
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

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 Cérou [INRIA, Researcher]
- Cédric Herzet [INRIA, Researcher, HDR]
- Patrick Héas [INRIA, Researcher]
- François Le Gland [INRIA, Senior Researcher, (until Nov. 2022 - retirement)]

Faculty Member

- Valérie Monbet [UNIV RENNES I, Professor, HDR]

PhD Students

- François Ernoult [UNIV RENNES I]
- Théo Guyard [INSA RENNES]
- Thu Le Tran [UNIV RENNES I]

Administrative Assistant

- Gunther Tessier [INRIA]

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 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. 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 MATHERIALS 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 New software and platforms

5.1 New software

5.1.1 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: Clément Elvira, Theo Guyard, Cedric Herzet

5.1.2 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: Clément Elvira, Theo Guyard, Cedric Herzet

5.1.3 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

Participants: Valérie Monbet, Thi Tuyet Trang Chau

5.1.4 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

Participant: Valérie Monbet

5.1.5 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`

Contact: Patrick Heas

Participant: Patrick Heas

5.1.6 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

Participants: Clément Elvira, Cedric Herzet

6 New results

6.1 Objective 1 – Rare events and Monte Carlo simulation

Monte-Carlo simulation

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

In [21], we obtained the first large deviation analysis of the (large sample size) statistical fluctuations of the AMS algorithm. The obtained limiting quantity can provide insights on the algorithmic efficiency in practice, in particular a novel geometric criterion ensuring minimal fluctuations (asymptotic efficiency) is studied.

In [22] we study a real world high dimensional Bayesian sampling problem (weather variables observed by space imagery) using kinetic Langevin diffusions (Hamiltonian Monte Carlo), and show empirically the advantage for convergence of an artificial "cold" tempering taming the non-linearities of the likelihood.

In [20], we obtained a theoretical result on Importance Sampling that proves the following fact: Consider a convex set of possible target probability distributions and a reference measure. The Gibbs-like distribution that minimizes entropy (with respect to the reference) on the considered convex class is also, in some rigorously defined worst-case sense, the optimal importance proposal.

In [8], we introduce and study a new family of simulable velocity jump Markov process (PDMP) with prescribed (up to normalization constant) stationary distribution (no time step error nor Metropolis correction !) that can converge towards kinetic Langevin diffusions.

6.2 Objective 2 – New topics in particle filtering and semi-parametric statistics

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

Model selection of climate and weather prediction models is a critical issue which can be tackled by constructing so-called analog forecasts; which are cheap stochastic generators of the output of different models constructed using historical simulation data. The latter can be combined with state-of-art Monte Carlo filtering procedures (e.g. Gaussian-based Ensemble Kalman filters) to efficiently compare the likelihood of the prediction output of the considered different models evaluated on some *real in situ* observations. These applications have been studied in [26, 9].

The above results have motivated the development of an original semi-parametric inference methodology able to construct stochastic weather models/generators, the non-parametric part relying on a "catalog of analogs" consisting of past data (*e.g.* a time series). In [1], a hidden latent Markovian model with parametric noise and non-parametric drift is inferred from an historical catalog of data (a time series) using a stochastic Expectation-Maximisation/Estimation iterative scheme (with iteration index k). In the latter, the smoothed (conditional on data) distribution of the step k hidden model is simulated using advanced pathwise sequential Monte Carlo particle filters (Conditional Particle Filters with Backward simulation). The step $k+1$ model is then estimated using both parametric maximum likelihood and non-parametric / machine learning tools.

6.3 Objective 3 – Semi-parametric and applied statistics

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

Motivated by applications to weather/climate data, various original estimation methods have been proposed, for instance a new Expectation-Maximisation method for generalized ridge regression in [24]. Other works have compared various statistical methods optimally chosen and adjusted to time-series in meteorological contexts: regression ([25]) for wave height/wind relation; Gaussian mixtures for calibration of ensemble forecasts ([5]); regression and Deep/Machine Learning ([7, 23]) for the down-scaling of sea states. In [6] stochastic weather generators developed and studied by the team are used for the design and reliability evaluation of desalination systems.

In [10], prediction of allergic pollen risk from meteorological data and assimilation is studied.

6.4 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. [4] 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 ℓ^2-norm of the optimal approximation error.

Another issue occurs when using prior model to solve reconstruction tasks where one wants to recover some quantity of interest from partial/noisy observations. In many situations, given the inputs of the problem at hand, some parts of the model may be irrelevant to solve the target reconstruction task. Hence, a recent trend (e.g. for "k-sparse" models) consists in the "on-line" simplification of the prior model, "on-line" meaning here "*during* the reconstruction process". This approach (named "screening" in the case of k-sparse models) can thus be seen to some extent as a technique for "online model reduction" and aims to achieve better accuracy/complexity trade-offs. We showed that the principle of screening goes well beyond the standard k-sparse model.We were the first team to derive a valid screening method applying on "non-separable" sparse models, see [3]; see also [15] and applications to ℓ_1 or LASSO models in [11, 13, 17, 16]. We also have investigated how the mechanics of screening can be extended to more general problems involving non-convex functions: in [12, 14] we dealt with the identification of zeros in the solution of optimization problems involving the "ℓ_0" counting function.

Finally, in [2], we study the design of sensor arrays in the context involving the localization of a few acoustic sources using sparse approximation to find the source locations.

7 Bilateral contracts and grants with industry

7.1 Bilateral contracts with industry

7.1.1 CIFRE grants

- **LIFY air**.

  Participants: Valérie Monbet.

  Through the CIFRE PhD project of Esso-Ridah Bleza supervised by Valérie Monbet on IA for multi-source pollen detection, see [10].

7.1.2 Meteorological Satellite Data Processing

Participants: Patrick Héas.

Industrial Partner: EUMETSAT of Darmstadt.

Partner Contact:`Regis.Borde@eumetsat.int`

The transferred technology concerns an algorithm for the operational and real-time production of vertically resolved 3D atmospheric motion vector fields (AMVs) from measurements of new hyperspectral instruments: the infrared radiosounders on the third generation Meteosat satellites (MTG), developed by the European Space Agency (ESA) and the Infrared Atmospheric Sounding Interferometer (IASI) on MetOp-A and MetOp-B developed by the French Space Agency (CNES).

8 Partnerships and cooperations

8.1 International initiatives

8.1.1 Participation in other International Programs

Participants: Valérie Monbet.

Title: ECOS ARGENTINE

ECOS ARGENTINE. Funding program through the ECOS Sud - MINCyT intiative. The program involves a collaboration with the French-Argentinian Climate Institute.

8.2 National initiatives

8.2.1 ANR

ANR MELODY (2020-2024)

Participants: Cédric Herzet.

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).

The partners involved in the project were: IMT Atlantique (PI: Ronan Fablet), Inria-Rennes, Inria-Grenoble, Laboratoire d'Océanographie Physique et Spatiale, Institut des géosciences et de l'environnement, Institut Pierre-Simon Laplace.

ANR SINEQ (2021-2025)

Participants: Mathias Rousset, Frédéric Cérou.

Simulating non-equilibrium stochastic dynamics. The goal of the SINEQ project is, within a mathematical perspective, to extend various variance reduction techniques used in the Monte Carlo computation of equilibrium properties of statistical physics models.

The partners involved in the project are: CERMICS (PI: G. Stoltz), CEREMADE and Inria Rennes.

9 Dissemination

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

9.1 PhD and HDR defenses

Cédric Herzet has defended his HdR (Habilitation à Diriger des Recherches) on november 24, 2022. Manuscript: [19].

Audrey Cuillery has defended her PhD in 2022. Manuscript: [18] (supervision: F. Le Gland).

Saïd Obakrim has defended his PhD in 2022. (supervision: V. Monbet).

9.2 Promoting scientific activities

9.2.1 Scientific events: organisation

F. Cérou is organizing the weekly Seminar 'Stochastic Processes' at IRMAR, Univ Rennes.

F. Cérou has organized a special session at MCQMC 2018 in Rennes, and has organized a special session and given a talk at MCQMC 2022 in Linz (Austria).

V. Monbet has co-organized the Workshop "Machine learning and uncertainties in climate simulations", within the Thematic Semester 2022 of Labex Lebesgue.

M. Rousset has co-organized the Workshop Probabilistic Numerical Methods for Machine Learning: recent trends (1-3 June, Rennes) within the Thematic Semester 2022 of Labex Lebesgue.

9.2.2 Journal

Reviewer - reviewing activities Many, for various journals (e.g. probability: EJP, AAP ...; signal processing: various IEE, ...; statistics: JRSS (C), ...).

9.2.3 Invited talks

Some examples:

P. Héas, Model Reduction and Surrogate Modeling (MORE), Berlin, Germany, 2022.

F. Cérou has given a talk at MCQMC 2022 in Linz (Austria).

C. Herzet , ICASSP 2022 - IEEE International Conference on Acoustic, Speech and Signal Processing.

C. Herzet,2022 : European Signal Processing Conference.

9.3 Teaching - Supervision - Juries

9.3.1 Teaching

Cédric Herzet has given:

- Algorithmique et programmation, ENSAI, Licence 3 : Responsable du module, 6h CM, 3h TD, 12 TP
- Traitement du signal, ENS Rennes, Master 1 : Responsable du module, 6h CM
- Représentations parcimonieuses, INSA Rennes, Master 2 : Responsable du module, 6hCM
- Régularisation, ENSAI, Master 2 : Responsable du module, 6h CM, 12TP
- Encadrement projet recherche, INSA, Master 1 : 10h
- Gestion de la filière "Data Science et Ingénierie des Données", Master 2, ENSAI : 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 4h specialized course 'Large Deviations Theory' and another 4h specialized 'distcretization of SDEs' for PhD students program, Univ Rennes.

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

9.3.2 Supervision

M. Rousset has been supervising the PhD of: Karim Tit: *Rare event analysis of the Reliability of Deep Neural Networks*, CIFRE Inria and Thalès, starting Jan. 2021, co-supervision: T.Furon.

M. Rousset has been supervising the PhD of François Ernoult, *Small noise analysis of rare event splitting algorithms*, UR1 and Région Bretagne, starting Sept. 2020.

V. Monbet has been supervising the PhD of: Said Obakrim, *Statistical downscaling and climate change in the coastal zone.* UR1, defended Oct. 2022. Co-supervision: Pierre Ailliot.

C. Herzet has been supervising the PhD of Théo Guyard, *Screening methods for non-convex sparse representations.* INSA, starting Sept. 2021, co-supervision James Ledoux.

C. Herzet has been supervising the PhD of Le Thu Tran, *Sparse representations in continuous dictionaries. Application to spectroscopy data.* UR1, sarting Sept. 2020, co-supervision: Valérie Monbet.

F. Le Gland has been supervising the PhD of Audrey Cuillery [18]. CIFRE Inria and Naval group, defended June 2022.

9.3.3 Juries

V. Monbet has been member of the following PhD defense committees:

- J. Legrand (LSCE) Simulation and assessment of multivariate extreme models for environmental data
- B. Balogh (MeteoFrance) Vers une utilisation de l'Intelligence Artificielle dans un modèle numérique de climat (reviewer)
- A. Troncosa (Mines Paris Tech) Conditional simulations of reservoir models using Sequential Monte-Carlo methods (reviewer)

V. Monbet has been member of the following HdR defense committees:

- C. Herzet (INRIA), Some contributions to reconstruction problems.
- M. Marbac-Loudelle (ENSAI), Méthodes de classifications.

10 Scientific production

10.1 Publications of the year

International journals

[1] T. T. T. Chau, P. Ailliot, V. Monbet and P. Tandeo. 'Comparison of simulation-based algorithms for parameter estimation and state reconstruction in nonlinear state-space models'. In: *Discrete and Continuous Dynamical Systems - Series S* (2022), pp. 1–24. DOI: 10.3934/dcdss.2022054. URL: https://imt-atlantique.hal.science/hal-03616079.

[2] M. Courcoux-Caro, C. Vanwynsberghe, C. Herzet and A. Baussard. 'Sequential sensor selection for the localization of acoustic sources by sparse Bayesian learning'. In: *Journal of the Acoustical Society of America* 152.3 (19th Sept. 2022), pp. 1695–1708. DOI: 10.1121/10.0014001. URL: https://hal.archives-ouvertes.fr/hal-03780626.

[3] C. Elvira and C. Herzet. 'Safe rules for the identification of zeros in the solution of the SLOPE problem'. In: *SIAM Journal on Mathematics of Data Science* (2022). URL: https://hal.archives-ouvertes.fr/hal-03400322.

[4] P. Héas and C. Herzet. 'Low-Rank Dynamic Mode Decomposition: An Exact and Tractable Solution'. In: *Journal of Nonlinear Science* 32.1 (Feb. 2022), article n°8. DOI: 10.1007/s00332-021-09770-w. URL: https://hal.inria.fr/hal-03468966.

[5] G. Jouan, A. Cuzol, V. Monbet and G. Monnier. 'Gaussian mixture models for clustering and calibration of ensemble weather forecasts'. In: *Discrete and Continuous Dynamical Systems - Series S* (2022), pp. 1–19. DOI: 10.3934/dcdss.2022037. URL: https://hal.science/hal-03619364.

[6] E. Koutroulis, G. Petrakis, V. Agou, A. Malisovas, D. Hristopulos, P. Partsinevelos, A. Tripolitsiotis, N. Halouani, P. Ailliot, M. Boutigny, V. Monbet, D. Allard, A. Cuzol, D. Kolokotsa, E. Varouchakis, K. Kokolakis and S. Mertikas. 'Site selection and system sizing of desalination plants powered with renewable energy sources based on a web-GIS platform'. In: *International Journal of Energy Sector Management* 16.3 (25th Mar. 2022), pp. 469–492. DOI: 10.1108/IJESM-04-2021-0018. URL: https://hal.science/hal-03378232.

[7] M. Michel, S. Obakrim, N. Raillard, P. Ailliot and V. Monbet. 'Deep learning for statistical downscaling of sea states'. In: *Advances in Statistical Climatology, Meteorology and Oceanography* 8 (7th Apr. 2022), pp. 83–95. DOI: 10.5194/ascmo-8-83-2022. URL: https://hal.science/hal-03825410.

[8] P. Monmarché, M. Rousset and P.-A. Zitt. 'Exact targeting of Gibbs distributions using velocity-jump processes'. In: *Stochastics and Partial Differential Equations: Analysis and Computations* (2022). DOI: 10.1007/s40072-022-00247-9. URL: https://hal.archives-ouvertes.fr/hal-02916073.

[9] J. Ruiz, P. Ailliot, T. Tuyet Trang Chau, P. Le Bras, V. Monbet, F. Sévellec and P. Tandeo. 'Analog data assimilation for the selection of suitable general circulation models'. In: *Geoscientific Model Development* 15 (2022), pp. 7203–7220. DOI: `10.5194/gmd-15-7203-2022`. URL: `https://hal-insu.archives-ouvertes.fr/insu-03868833`.

Conferences without proceedings

[10] E.-R. Bleza, V. Monbet and P.-F. Marteau. 'Prédiction des niveaux de risque pollinique à partir de données historiques multi-sources'. In: EGC 2022 - Extraction et Gestion des Connaissances. Blois, France, 24th Jan. 2022. URL: `https://hal.science/hal-03614274`.

[11] T. Guyard, C. Herzet and C. Elvira. 'Screen & Relax: Accelerating The Resolution Of Elastic-Net By Safe Identification of The Solution Support'. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore: IEEE, 23rd May 2022, pp. 5443–5447. DOI: `10.1109/ICASSP43922.2022.9747412`. URL: `https://hal-univ-rennes1.archives-ouvertes.fr/hal-03688032`.

[12] T. Guyard, C. Herzet and C. Elvira. 'Node-Screening Tests For The L0-Penalized Least-Squares Problem'. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore: IEEE, 23rd May 2022, pp. 5448–5452. DOI: `10.1109/ICASSP43922.2022.9747563`. URL: `https://hal-univ-rennes1.archives-ouvertes.fr/hal-03688011`.

[13] T. Guyard, C. Herzet and C. Elvira. 'Screen & Relax: Accélérer la résolution du problème "Elastic-Net" par identification du support de la solution'. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 5th Sept. 2022, pp. 1–4. URL: `https://hal.archives-ouvertes.fr/hal-03778139`.

[14] T. Guyard, C. Herzet, C. Elvira and A. N. Arslan. 'Node-screening pour le problème des moindres carrés avec pénalité L0'. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 6th Sept. 2022, pp. 1–4. URL: `https://hal.archives-ouvertes.fr/hal-03784682`.

[15] C. Herzet, C. Elvira and H.-P. Dang. 'Region-free Safe Screening Tests for l1-penalized Convex Problems'. In: Eusipco 2022 - 30th European Signal Processing Conference. Belgrade, Serbia, 29th Aug. 2022, pp. 1–5. URL: `https://hal-centralesupelec.archives-ouvertes.fr/hal-03806099`.

[16] T.-L. Tran, C. Elvira, H.-P. Dang and C. Herzet. 'Beyond GAP screening for Lasso by exploiting new dual cutting half-spaces with supplementary material'. In: Eusipco 2022 - 30th European Signal Processing Conference. Vol. 2022. Belgrade, Serbia, 29th Aug. 2022, pp. 2056–2060. URL: `https://hal-centralesupelec.archives-ouvertes.fr/hal-03805966`.

[17] T.-L. Tran, C. Elvira, H.-P. Dang and C. Herzet. 'Une nouvelle méthode d'accélération pour LASSO par élimination sûre de variables'. In: CAP 2022 - Conférence sur l'Apprentissage automatique. Vannes, France, 5th July 2022, pp. 1–6. URL: `https://hal-centralesupelec.archives-ouvertes.fr/hal-03806044`.

Doctoral dissertations and habilitation theses

[18] A. Cuillery. 'Bayesian multitarget tracking from raw data'. Université Rennes 1, 17th June 2022. URL: `https://theses.hal.science/tel-03850485`.

[19] C. Herzet. 'Some Contributions to Reconstruction Problems'. Université de Rennes 1, 24th Nov. 2022. URL: `https://hal.inria.fr/tel-03888741`.

Reports & preprints

[20] F. Cérou, P. Héas and M. Rousset. *Entropy minimizing distributions are worst-case optimal importance proposals.* 8th Dec. 2022. URL: `https://hal.archives-ouvertes.fr/hal-03889404`.

[21] F. Cérou, S. Martel and M. Rousset. *Fluctuations of Rare Event Simulation with Monte Carlo Splitting in the Small Noise Asymptotics*. 8th Dec. 2022. URL: `https://hal.archives-ouvertes.fr/hal-03889692`.

[22] P. Héas, F. Cérou and M. Rousset. *Uncertainty of Atmospheric Motion Vectors by Sampling Tempered Posterior Distributions*. 15th Sept. 2022. URL: `https://hal.inria.fr/hal-03777922`.

[23] V. Monbet, S. Obakrim, P. Ailliot and N. Raillard. *LEARNING THE SPATIO-TEMPORAL RELATIONSHIP BETWEEN WIND AND SIGNIFICANT WAVE HEIGHT USING DEEP LEARNING*. 22nd Oct. 2022. URL: `https://hal.science/hal-03825412`.

[24] V. Monbet, S. Obakrim, N. Raillard and P. Ailliot. *Em algorithm for generalized ridge regression with spatial covariates*. 22nd Oct. 2022. URL: `https://hal.science/hal-03825411`.

[25] S. Obakrim, P. Ailliot, V. Monbet and N. Raillard. *Statistical modeling of the space-time relation between wind and significant wave height*. 22nd Oct. 2022. DOI: `10.1002/essoar.10510147.2`. URL: `https://hal.science/hal-03825413`.

[26] J. Ruiz, P. Ailliot, P. Le Bras, V. Monbet, F. Sévellec, T. T. T. Chau and P. Tandeo. *Analog Data Assimilation for the Selection of Suitable General Circulation Models*. 2022. DOI: `10.5194/gmd-2021-434`. URL: `https://imt-atlantique.hal.science/hal-03685531`.