Activity Report 2017

Team ASPI

Applications of interacting particle systems to statistics

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

1. Personnel 1
2. Overall Objectives 2
3. Research Program 3
3.1. Interacting Monte Carlo methods and particle approximation of Feynman–Kac distributions 3
3.2. Multilevel splitting for rare event simulation 4
3.3. Statistical learning: pattern recognition and nonparametric regression 6
4. Application Domains 7
4.1. Localisation, navigation and tracking 7
4.2. Rare event simulation 7
5. New Results 8
5.1. Central limit theorem for adaptive multilevel splitting 8
5.2. Adaptive multilevel splitting for Monte Carlo particle transport 8
5.3. Weak overdamped limit theorem for Langevin processes 8
5.4. Particle–Kalman filter for structural health monitoring 8
5.5. Reduced modeling of unknown trajectories 9
5.6. Model reduction from partial observations 9
5.7. Low–rank dynamic mode decomposition: optimal solution in polynomial time 10
5.8. Optimal kernel–based dynamic mode decomposition 10
5.9. Non parametric state–space model for missing–data imputation 10
6. Bilateral Contracts and Grants with Industry 11
6.1.1. Hybrid indoor navigation — PhD project at CEA LETI 11
6.1.2. Bayesian tracking from raw data — CIFRE grant with DCNS Nantes 11
7. Partnerships and Cooperations 12
7.1. Regional initiatives 12
7.2. National initiatives 12
7.2.1. Computational Statistics and Molecular Simulation (COSMOS) — ANR challenge Information and Communication Society 12
7.2.2. Advanced Geophysical Reduced–Order Model Construction from Image Observations (GERONIMO) — ANR programme Jeunes Chercheuses et Jeunes Chercheurs 13
7.3. European initiatives 13
7.3.1. Molecular Simulation: Modeling, Algorithms and Mathematical Analysis (MSMaths) — ERC Consolidator Grant 13
7.3.2. Design of Desalination Systems Based on Optimal Usage of Multiple Renewable Energy Sources (DESIRES) — ERANETMED NEXUS–14–049 13
7.4. International initiatives 14
7.5. International research visitors 14
8. Dissemination 14
8.1. Promoting scientific activities 14
8.1.1. Scientific events organisation 14
8.1.2. Participation in workshops, seminars, lectures, etc. 14
8.1.3. Research administration 15
8.2. Teaching, supervision, thesis committees 15
8.2.1. Teaching 15
8.2.2. Supervision 16
8.2.3. Thesis committees 16
9. Bibliography 17

Team ASPI

Creation of the Project-Team: 2005 January 10, updated into Team: 2017 January 01, end of the Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A3.4.5. - Bayesian methods
A3.4.7. - Kernel methods
A5.9.2. - Estimation, modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling (SPDE, SDE)
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.3.2. - Data assimilation
A6.3.4. - Model reduction

Other Research Topics and Application Domains:

B2.3. - Epidemiology
B3.2. - Climate and meteorology
B3.3.2. - Water: sea & ocean, lake & river
B3.3.4. - Atmosphere
B7.1.3. - Air traffic
B9.4.3. - Physics
B9.4.4. - Chemistry

1. Personnel

Research Scientists

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

Faculty Member

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

Post-Doctoral Fellow

Hassan Maatouk [Inria, until Aug 2017]

PhD Students

Thi Tuyet Trang Chau [Univ. Rennes 1]
Ramatoulaye Dabo [Univ. Assane Seck de Ziguinchor and Univ. Rennes 1, co–tutelle]
Audrey Cuillery [Naval Group]
Kersane Zoubert–Ousseni [CEA, until Nov 2017]

Administrative Assistant

Fabienne Cuyollaa [Inria]

External Collaborator

Arnaud Guyader [Univ Pierre et Marie Curie, HDR]

2. Overall Objectives

2.1. Overall Objectives

The scientific objectives of ASPI are the design, analysis and implementation of interacting Monte Carlo methods, also known as particle methods, with focus on

- statistical inference in hidden Markov models and particle filtering,
- risk evaluation and simulation of rare events,
- global optimization.

The whole problematic is multidisciplinary, not only because of the many scientific and engineering areas in which particle methods are used, but also because of the diversity of the scientific communities which have already contributed to establish the foundations of the field

> target tracking, interacting particle systems, empirical processes, genetic algorithms (GA), hidden Markov models and nonlinear filtering, Bayesian statistics, Markov chain Monte Carlo (MCMC) methods, etc.

Intuitively speaking, interacting Monte Carlo methods are sequential simulation methods, in which particles

- *explore* the state space by mimicking the evolution of an underlying random process,
- *learn* their environment by evaluating a fitness function,
- and *interact* so that only the most successful particles (in view of the fitness function) are allowed to survive and to get offsprings at the next generation.

The effect of this mutation / selection mechanism is to automatically concentrate particles (i.e. the available computing power) in regions of interest of the state space. In the special case of particle filtering, which has numerous applications under the generic heading of positioning, navigation and tracking, in

> target tracking, computer vision, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.,

each particle represents a possible hidden state, and is replicated or terminated at the next generation on the basis of its consistency with the current observation, as quantified by the likelihood function. With these genetic–type algorithms, it becomes easy to efficiently combine a prior model of displacement with or without constraints, sensor–based measurements, and a base of reference measurements, for example in the form of a digital map (digital elevation map, attenuation map, etc.). In the most general case, particle methods provide approximations of Feynman–Kac distributions, a pathwise generalization of Gibbs–Boltzmann distributions, by means of the weighted empirical probability distribution associated with an interacting particle system, with applications that go far beyond filtering, in

> simulation of rare events, global optimization, molecular simulation, etc.

The main applications currently considered are geolocalisation and tracking of mobile terminals, terrain–aided navigation, data fusion for indoor localisation, optimization of sensors location and activation, risk assessment in air traffic management, protection of digital documents.

3. Research Program

3.1. Interacting Monte Carlo methods and particle approximation of Feynman–Kac distributions

Monte Carlo methods are numerical methods that are widely used in situations where (i) a stochastic (usually Markovian) model is given for some underlying process, and (ii) some quantity of interest should be evaluated, that can be expressed in terms of the expected value of a functional of the process trajectory, which includes as an important special case the probability that a given event has occurred. Numerous examples can be found, e.g. in financial engineering (pricing of options and derivative securities) [36], in performance evaluation of communication networks (probability of buffer overflow), in statistics of hidden Markov models (state estimation, evaluation of contrast and score functions), etc. Very often in practice, no analytical expression is available for the quantity of interest, but it is possible to simulate trajectories of the underlying process. The idea behind Monte Carlo methods is to generate independent trajectories of this process or of an alternate instrumental process, and to build an approximation (estimator) of the quantity of interest in terms of the weighted empirical probability distribution associated with the resulting independent sample. By the law of large numbers, the above estimator converges as the size N of the sample goes to infinity, with rate $1/\sqrt{N}$ and the asymptotic variance can be estimated using an appropriate central limit theorem. To reduce the variance of the estimator, many variance reduction techniques have been proposed. Still, running independent Monte Carlo simulations can lead to very poor results, because trajectories are generated *blindly*, and only afterwards are the corresponding weights evaluated. Some of the weights can happen to be negligible, in which case the corresponding trajectories are not going to contribute to the estimator, i.e. computing power has been wasted.

A major breakthrough made in the mid 90's, has been the introduction of interacting Monte Carlo methods, also known as sequential Monte Carlo (SMC) methods, in which a whole (possibly weighted) sample, called *system of particles*, is propagated in time, where the particles

- *explore* the state space under the effect of a *mutation* mechanism which mimics the evolution of the underlying process,
- and are *replicated* or *terminated*, under the effect of a *selection* mechanism which automatically concentrates the particles, i.e. the available computing power, into regions of interest of the state space.

In full generality, the underlying process is a discrete–time Markov chain, whose state space can be

> finite, continuous, hybrid (continuous / discrete), graphical, constrained, time varying, pathwise, etc.,

the only condition being that it can easily be *simulated*.

In the special case of particle filtering, originally developed within the tracking community, the algorithms yield a numerical approximation of the optimal Bayesian filter, i.e. of the conditional probability distribution of the hidden state given the past observations, as a (possibly weighted) empirical probability distribution of the system of particles. In its simplest version, introduced in several different scientific communities under the name of *bootstrap filter* [38], *Monte Carlo filter* [43] or *condensation* (conditional density propagation) algorithm [42], and which historically has been the first algorithm to include a resampling step, the selection mechanism is governed by the likelihood function: at each time step, a particle is more likely to survive and to replicate at the next generation if it is consistent with the current observation. The algorithms also provide as a by–product a numerical approximation of the likelihood function, and of many other contrast functions for parameter estimation in hidden Markov models, such as the prediction error or the conditional least–squares criterion.

Particle methods are currently being used in many scientific and engineering areas

> positioning, navigation, and tracking [39], [32], visual tracking [42], mobile robotics [33], [55], ubiquitous computing and ambient intelligence, sensor networks, risk evaluation and simulation of rare events [37], genetics, molecular simulation [34], etc.

Other examples of the many applications of particle filtering can be found in the contributed volume [22] and in the special issue of *IEEE Transactions on Signal Processing* devoted to *Monte Carlo Methods for Statistical Signal Processing* in February 2002, where the tutorial paper [23] can be found, and in the textbook [51] devoted to applications in target tracking. Applications of sequential Monte Carlo methods to other areas, beyond signal and image processing, e.g. to genetics, can be found in [48]. A recent overview can also be found in [25].

Particle methods are very easy to implement, since it is sufficient in principle to simulate independent trajectories of the underlying process. The whole problematic is multidisciplinary, not only because of the already mentioned diversity of the scientific and engineering areas in which particle methods are used, but also because of the diversity of the scientific communities which have contributed to establish the foundations of the field

> target tracking, interacting particle systems, empirical processes, genetic algorithms (GA), hidden Markov models and nonlinear filtering, Bayesian statistics, Markov chain Monte Carlo (MCMC) methods.

These algorithms can be interpreted as numerical approximation schemes for Feynman–Kac distributions, a pathwise generalization of Gibbs–Boltzmann distributions, in terms of the weighted empirical probability distribution associated with a system of particles. This abstract point of view [30], [29], has proved to be extremely fruitful in providing a very general framework to the design and analysis of numerical approximation schemes, based on systems of branching and / or interacting particles, for nonlinear dynamical systems with values in the space of probability distributions, associated with Feynman–Kac distributions. Many asymptotic results have been proved as the number N of particles (sample size) goes to infinity, using techniques coming from applied probability (interacting particle systems, empirical processes [58]), see e.g. the survey article [30] or the textbooks [29], [28], and references therein

> convergence in $\mathbb{L}^p$, convergence as empirical processes indexed by classes of functions, uniform convergence in time, see also [46], [47], central limit theorem, see also [44], [31], propagation of chaos, large deviations principle, etc.

The objective here is to systematically study the impact of the many algorithmic variants on the convergence results.

3.2. Multilevel splitting for rare event simulation

See 4.2, 5.1, and 5.2.

The estimation of the small probability of a rare but critical event, is a crucial issue in industrial areas such as

> nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [45], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. As shown in [6], the Feynman–Kac formalism of 3.1 is well suited for the design and analysis of splitting algorithms for rare event simulation.

Propagation of uncertainty Multilevel splitting can be used in static situations. Here, the objective is to learn the probability distribution of an output random variable $Y = F(X)$, where the function F is only defined pointwise for instance by a computer programme, and where the probability distribution of the input random variable X is known and easy to simulate from. More specifically, the objective could be to compute the probability of the output random variable exceeding a threshold, or more generally to evaluate the cumulative distribution function of the output random variable for different output values. This problem is characterized by the lack of an analytical expression for the function, the computational cost of a single pointwise evaluation of the function, which means that the number of calls to the function should be limited as much as possible, and finally the complexity and / or unavailability of the source code of the computer programme, which makes any modification very difficult or even impossible, for instance to change the model as in importance sampling methods.

The key issue is to learn as fast as possible regions of the input space which contribute most to the computation of the target quantity. The proposed splitting methods consists in (i) introducing a sequence of intermediate regions in the input space, implicitly defined by exceeding an increasing sequence of thresholds or levels, (ii) counting the fraction of samples that reach a level given that the previous level has been reached already, and (iii) improving the diversity of the selected samples, usually with an artificial Markovian dynamics for the input variable. In this way, the algorithm learns

- the transition probability between successive levels, hence the probability of reaching each intermediate level,
- and the probability distribution of the input random variable, conditionned on the output variable reaching each intermediate level.

A further remark, is that this conditional probability distribution is precisely the optimal (zero variance) importance distribution needed to compute the probability of reaching the considered intermediate level.

Rare event simulation To be specific, consider a complex dynamical system modelled as a Markov process, whose state can possibly contain continuous components and finite components (mode, regime, etc.), and the objective is to compute the probability, hopefully very small, that a critical region of the state space is reached by the Markov process before a final time T, which can be deterministic and fixed, or random (for instance the time of return to a recurrent set, corresponding to a nominal behaviour).

The proposed splitting method consists in (i) introducing a decreasing sequence of intermediate, more and more critical, regions in the state space, (ii) counting the fraction of trajectories that reach an intermediate region before time T, given that the previous intermediate region has been reached before time T, and (iii) regenerating the population at each stage, through resampling. In addition to the non–intrusive behaviour of the method, the splitting methods make it possible to learn the probability distribution of typical critical trajectories, which reach the critical region before final time T, an important feature that methods based on importance sampling usually miss. Many variants have been proposed, whether

- the branching rate (number of offsprings allocated to a successful trajectory) is fixed, which allows for depth–first exploration of the branching tree, but raises the issue of controlling the population size,
- the population size is fixed, which requires a breadth–first exploration of the branching tree, with random (multinomial) or deterministic allocation of offsprings, etc.

Just as in the static case, the algorithm learns

- the transition probability between successive levels, hence the probability of reaching each intermediate level,
- and the entrance probability distribution of the Markov process in each intermediate region.

Contributions have been given to

- minimizing the asymptotic variance, obtained through a central limit theorem, with respect to the shape of the intermediate regions (selection of the importance function), to the thresholds (levels), to the population size, etc.
- controlling the probability of extinction (when not even one trajectory reaches the next intermediate level),
- designing and studying variants suited for hybrid state space (resampling per mode, marginalization, mode aggregation),

and in the static case, to

- minimizing the asymptotic variance, obtained through a central limit theorem, with respect to intermediate levels, to the Metropolis kernel introduced in the mutation step, etc.

A related issue is global optimization. Indeed, the difficult problem of finding the set M of global minima of a real–valued function V can be replaced by the apparently simpler problem of sampling a population from a probability distribution depending on a small parameter, and asymptotically supported by the set M as the small parameter goes to zero. The usual approach here is to use the cross–entropy method [52], [27], which relies on learning the optimal importance distribution within a prescribed parametric family. On the other hand, multilevel splitting methods could provide an alternate nonparametric approach to this problem.

3.3. Statistical learning: pattern recognition and nonparametric regression

In pattern recognition and statistical learning, also known as machine learning, nearest neighbor (NN) algorithms are amongst the simplest but also very powerful algorithms available. Basically, given a training set of data, i.e. an N–sample of i.i.d. object–feature pairs, with real–valued features, the question is how to generalize, that is how to guess the feature associated with any new object. To achieve this, one chooses some integer k smaller than N, and takes the mean–value of the k features associated with the k objects that are nearest to the new object, for some given metric.

In general, there is no way to guess exactly the value of the feature associated with the new object, and the minimal error that can be done is that of the Bayes estimator, which cannot be computed by lack of knowledge of the distribution of the object–feature pair, but the Bayes estimator can be useful to characterize the strength of the method. So the best that can be expected is that the NN estimator converges, say when the sample size N grows, to the Bayes estimator. This is what has been proved in great generality by Stone [53] for the mean square convergence, provided that the object is a finite–dimensional random variable, the feature is a square–integrable random variable, and the ratio k/N goes to 0. Nearest neighbor estimator is not the only local averaging estimator with this property, but it is arguably the simplest.

The asymptotic behavior when the sample size grows is well understood in finite dimension, but the situation is radically different in general infinite dimensional spaces, when the objects to be classified are functions, images, etc.

Nearest neighbor classification in infinite dimension In finite dimension, the k–nearest neighbor classifier is universally consistent, i.e. its probability of error converges to the Bayes risk as N goes to infinity, whatever the joint probability distribution of the pair, provided that the ratio k/N goes to zero. Unfortunately, this result is no longer valid in general metric spaces, and the objective is to find out reasonable sufficient conditions for the weak consistency to hold. Even in finite dimension, there are exotic distances such that the nearest neighbor does not even get closer (in the sense of the distance) to the point of interest, and the state space needs to be complete for the metric, which is the first condition. Some regularity on the regression function is required next. Clearly, continuity is too strong because it is not required in finite dimension, and a weaker form of regularity is assumed. The following consistency result has been obtained: if the metric space is separable and if some Besicovich condition holds, then the nearest neighbor classifier is weakly consistent. Note that the Besicovich condition is always fulfilled in finite dimensional vector spaces (this result is called the Besicovich theorem), and that a counterexample [4] can be given in an infinite dimensional space with a Gaussian measure (in this case, the nearest neighbor classifier is clearly nonconsistent). Finally, a simple example has been found which verifies the Besicovich condition with a noncontinuous regression function.

Rates of convergence of the functional k–nearest neighbor estimator Motivated by a broad range of potential applications, such as regression on curves, rates of convergence of the k–nearest neighbor estimator of the regression function, based on N independent copies of the object–feature pair, have been investigated when the object is in a suitable ball in some functional space. Using compact embedding theory, explicit and general finite sample bounds can be obtained for the expected squared difference between the k–nearest neighbor estimator and the Bayes regression function, in a very general setting. The results have also been particularized to classical function spaces such as Sobolev spaces, Besov spaces and reproducing kernel Hilbert spaces. The rates obtained are genuine nonparametric convergence rates, and up to our knowledge the first of their kind for k–nearest neighbor regression.

This topic has produced several theoretical advances [1], [2] in collaboration with Gérard Biau (université Pierre et Marie Curie). A few possible target application domains have been identified in

- the statistical analysis of recommendation systems,
- the design of reduced–order models and analog samplers,

that would be a source of interesting problems.

4. Application Domains

4.1. Localisation, navigation and tracking

See 6.1.

Among the many application domains of particle methods, or interacting Monte Carlo methods, ASPI has decided to focus on applications in localisation (or positioning), navigation and tracking [39], [32], which already covers a very broad spectrum of application domains. The objective here is to estimate the position (and also velocity, attitude, etc.) of a mobile object, from the combination of different sources of information, including

- a prior dynamical model of typical evolutions of the mobile, such as inertial estimates and prior model for inertial errors,
- measurements provided by sensors,
- and possibly a digital map providing some useful feature (terrain altitude, power attenuation, etc.) at each possible position.

In some applications, another useful source of information is provided by

- a map of constrained admissible displacements, for instance in the form of an indoor building map,

which particle methods can easily handle (map-matching). This Bayesian dynamical estimation problem is also called filtering, and its numerical implementation using particle methods, known as particle filtering, has been introduced by the target tracking community [38], [51], which has already contributed to many of the most interesting algorithmic improvements and is still very active, and has found applications in

> target tracking, integrated navigation, points and / or objects tracking in video sequences, mobile robotics, wireless communications, ubiquitous computing and ambient intelligence, sensor networks, etc.

ASPI is contributing (or has contributed recently) to several applications of particle filtering in positioning, navigation and tracking, such as geolocalisation and tracking in a wireless network, terrain–aided navigation, and data fusion for indoor localisation.

4.2. Rare event simulation

See 3.2, 5.1, and 5.2.

Another application domain of particle methods, or interacting Monte Carlo methods, that ASPI has decided to focus on is the estimation of the small probability of a rare but critical event, in complex dynamical systems. This is a crucial issue in industrial areas such as

> nuclear power plants, food industry, telecommunication networks, finance and insurance industry, air traffic management, etc.

In such complex systems, analytical methods cannot be used, and naive Monte Carlo methods are clearly unefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [45], where trajectories going towards the critical set are given offsprings, thus increasing the number of trajectories that eventually reach the critical set. This approach not only makes it possible to estimate the probability of the rare event, but also provides realizations of the random trajectory, given that it reaches the critical set, i.e. provides realizations of typical critical trajectories, an important feature that methods based on importance sampling usually miss.

ASPI is contributing (or has contributed recently) to several applications of multilevel splitting for rare event simulation, such as risk assessment in air traffic management, detection in sensor networks, and protection of digital documents.

5. New Results

5.1. Central limit theorem for adaptive multilevel splitting

Participants: Frédéric Cérou, Arnaud Guyader, Mathias Rousset.

See 3.2, and 4.2.

This is a collaboration with Bernard Delyon (université de Rennes 1).

Fleming–Viot type particle systems represent a classical way to approximate the distribution of a Markov process with killing, given that it is still alive at a final deterministic time. In this context, each particle evolves independently according to the law of the underlying Markov process until its killing, and then branches instantaneously on another randomly chosen particle. While the consistency of this algorithm in the large population limit has been recently studied in several articles, our purpose here is to prove central limit theorems under very general assumptions. For this, we only suppose that the particle system does not explode in finite time, and that the jump and killing times have atomless distributions. In particular, this includes the case of elliptic diffusions with hard killing.

5.2. Adaptive multilevel splitting for Monte Carlo particle transport

Participant: Mathias Rousset.

See 3.2, and 4.2.

Simulation of neutron transport with Monte Carlo methods is a central issue in order to assess the aging of french nucelar plants.

In [49], we propose an alternative version of the AMS (adaptive multilevel splitting) algortihm, adapted for the first time to the field of particle tranport. Within this context, it can be used to build an unbiased estimator of any quantity associated with particle tracks, such as flux, reaction rates or even non–Boltzmann tallies. Furthermore, the effciency of the AMS algorithm is shown not to be very sensitive to variations of its input parameters, which makes it capable of significant variance reduction without requiring extended user effort.

5.3. Weak overdamped limit theorem for Langevin processes

Participant: Mathias Rousset.

This is a collaboration with Pierre-André Zitt (université Paris Est Marne-la-Vallée).

The Langevin stochastic process is the main model used in molecular dynamics simulation, for instance for the simulation of reactive trajectories of bio-chemical systems with rare event techniques.

In [21], we prove convergence in distribution of Langevin processes in the overdamped diffusion 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.

5.4. Particle–Kalman filter for structural health monitoring

Participant: Frédéric Cérou.

This is a joint work with EPI I4S (Inria Rennes–Bretagne Atlantique).

Standard filtering techniques for structural parameter estimation assume that the input force either is known exactly 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, a novel algorithm is proposed to estimate the force as additional state in parallel to the system parameters. Two concurrent filters are employed for parameters and force respectively. For the parameters, interacting particle–Kalman filter is employed targeting systems with correlated noise. Alongside a second filter is employed to estimate the seismic force acting on the structure. The proposal is numerically validated on a sixteen degrees–of–freedom mass–spring–damper system. The estimation results confirm the applicability of the proposed algorithm.

In another work, the same approach has been used for varying system parameters with correlated state and observation noise. The idea is to nest a bank of linear KFs (Kalman filters) for state estimation within a PF (particle filter) environment that estimates the parameters. This facilitates employing relatively less expensive linear KF for linear state estimation problem while costly PF is employed only for parameter estimation. Additionally, the proposed algorithm also takes care of those systems for which system and measurement noises are not uncorrelated as it is commonly idealized in standard filtering algorithms. As an example, for mechanical systems under ambient vibration it happens when acceleration response is considered as measurement. Thus the process and measurement noise in these system descriptions are obviously correlated. For this, an improved description for the Kalman gain is developed. Further, to enhance the consistency of particle filtering based parameter estimation involving high dimensional parameter space, a new temporal evolution strategy for the particles is defined. This strategy aims at restricting the solution from diverging (up to the point of no return) because of an isolated event of infeasible estimation which is very much likely especially when dealing with high dimensional parameter space.

5.5. Reduced modeling of unknown trajectories

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique)

In [12], 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, and *(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 as proper orthogonal decomposition (POD) or dynamic mode decomposition (DMD), can be revisited and recast within the probabilistic framework considered in this work. We illustrate the performance of the proposed approach by numerical results obtained for a standard geophysical model.

5.6. Model reduction from partial observations

Participant: Patrick Héas.

This is a collaboration with Angélique Drémeau (ENSTA Bretagne, Brest) and Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique)

In [11], we deal with model-order reduction of parametric partial differential equations (PPDE). More specifically, we consider the problem of finding a good approximation subspace of the solution manifold of the PPDE when only partial information on the latter is available. We assume that two sources of information are available: *i)* a "rough" prior knowledge, taking the form of a manifold containing the target solution manifold, and *ii)* partial linear measurements of the solutions of the PPDE (the term partial refers to the fact that observation operator cannot be inverted). We provide and study several tools to derive good approximation subspaces from these two sources of information. We first identify the best worst-case performance achievable in this setup and propose simple procedures to approximate the corresponding optimal approximation subspace. 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.

5.7. Low–rank dynamic mode decomposition: optimal solution in polynomial time

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique)

The works [15] and [41] study the linear approximation of high–dimensional dynamical systems using low-rank dynamic mode decomposition (DMD). Searching this approximation in a data–driven approach can be formalised as attempting to solve a low-rank constrained optimisation problem. This problem is non–convex and state–of–the–art algorithms are all sub–optimal. We show that there exists a closed-form solution, which can be computed in polynomial time, and characterises the ℓ_2–norm of the optimal approximation error. The theoretical results serve to design low–complexity algorithms building reduced models from the optimal solution, based on singular value decomposition or low–rank DMD. The algorithms are evaluated by numerical simulations using synthetic and physical data benchmarks.

5.8. Optimal kernel–based dynamic mode decomposition

Participant: Patrick Héas.

This is a collaboration with Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique)

The state–of–the–art algorithm known as kernel-based dynamic mode decomposition (K–DMD) provides a sub–optimal solution to the problem of reduced modeling of a dynamical system based on a finite approximation of the Koopman operator. It relies on crude approximations and on restrictive assumptions. The purpose of the work in [20] is to propose a kernel–based algorithm solving exactly this low–rank approximation problem in a general setting.

5.9. Non parametric state–space model for missing–data imputation

Participants: Thi Tuyet Trang Chau, François Le Gland, Valérie Monbet, Mathias Rousset.

This is a collaboration with Pierre Ailliot (université de Bretagne Occidentale, Brest), Ronan Fablet and Pierre Tandéo (Télécom Bretagne, Brest), Anne Cuzol (université de Bretagne Sud, Vannes) and Bernard Chapron (IFREMER, Brest).

Missing data are present in many environmental data–sets and this work aims at developing a general method for imputing them. State–space models (SSM) have already extensively been used in this framework. The basic idea consists in introducing the true environmental process, which we aim at reconstructing, as a latent process and model the data available at neighboring sites in space and/or time conditionally to this latent process. A key input of SSMs is a stochastic model which describes the temporal evolution of the environmental process of interest. In many applications, the dynamic is complex and can hardly be described using a tractable parametric model. Here we investigate a data-driven method where the dynamical model is learned using a non-parametric approach and historical observations of the environmental process of interest. From a statistical point of view, we will address various aspects related to SSMs in a non–parametric framework. First we will discuss the estimation of the filtering and smoothing distributions, that is the distribution of the latent space given the observations, using sequential Monte Carlo approaches in conjunction with local linear regression. Then, a more difficult and original question consists in building a non–parametric estimate of the dynamics which takes into account the measurement errors which are present in historical data. We will propose an EM–like algorithm where the historical data are corrected recursively. The methodology will be illustrated and validated on an univariate toy example.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral grants with industry

See 4.1.

6.1.1. Hybrid indoor navigation — PhD project at CEA LETI

Participants: François Le Gland, Kersane Zoubert–Ousseni.

This is a collaboration with Christophe Villien (CEA LETI, Grenoble).

The issue here is user localization, and more generally localization–based services (LBS). This problem is addressed by GPS for outdoor applications, but no such general solution has been provided so far for indoor applications. The desired solution should rely on sensors that are already available on smartphones and other tablet computers. Inertial solutions that use MEMS (microelectromechanical system, such as accelerometer, magnetometer, gyroscope and barometer) are already studied at CEA. An increase in performance should be possible, provided these data are combined with other available data: map of the building, WiFi signal, modeling of perturbations of the magnetic field, etc. To be successful, advanced data fusion techniques should be used, such as particle filtering and the like, to take into account displacement constraints due to walls in the building, to manage several possible trajectories, and to deal with rather heterogeneous information (map, radio signals, sensor signals).

The main objective of this thesis is to design and tune localization algorithms that will be tested on platforms already available at CEA. Special attention is paid to particle smoothing and particle MCMC algorithms, to exploit some very precise information available at special time instants, e.g. when the user is clearly localized near a landmark point.

In some applications, real time estimation of the trajectory is not needed, and a post processing framework may provide a better estimation of this trajectory. In [57], we present and compare three different algorithms to improve a real time trajectory estimation. Actually, two different smoothing algorithms and the Viterbi algorithm are implemented and evaluated. These methods improve the regularity of the estimated trajectory by reducing switches between hypotheses.

Post processing indoor navigation is interesting, for example to develop crowdsourcing analysis. The post processing framework allows to provide a better estimation than in a real time framework. The main contribution of [17] is to present a piecewise parametrization using IMU (inertial measurement unit) and RSS (received signal strength) measurements only, which lead to an optimization problem. A Levenberg–Marquardt algorithm improved with simulated annealing and an adjustment of RSS measurements data leads to a good estimation (55% of the error less than 5 meters) of the trajectory.

6.1.2. Bayesian tracking from raw data — CIFRE grant with DCNS Nantes

Participants: François Le Gland, Audrey Cuillery.

This is a collaboration with Dann Laneuville (DCNS Nantes).

After the introduction of MHT (multi–hypothesis tracking) techniques in the nineties, multitarget tracking has recently seen promising developpments with the introduction of new algorithms such as the PHD (probability hypothesis density) filter [50], [56] or the HISP (hypothesised filter for independent stochastic populations) filter [40]. These techniques provide a unified multitarget model in a Bayesian framework [54], which makes it possible to design recursive estimators of a *multitarget probability density*. Two main approaches can be used here: sequential Monte Carlo (SMC, also kown as particle filtering), and Gaussian mixture (GM). A third approach, based on discretizing the state–space in a possibly adaptive way, could also be considered despite its larger computational load. These methods are well studied and provide quite good results for *contact output* data, which correspond to regularly spaced measurements of targets with a large SNR (signal–to–noise ratio). Here, the data is processed (compared with a detection threshold) in each resolution cell of the sensor, so as to provide a list of detections at a given time instant. Among these methods, the HISP filter has the best performance/computational cost ratio.

However, these classical methods are unefficient for targets with a low SNR, e.g. targets in far range or small targets with a small detection probability. For such targets, preprocessing (thresholding) the data is not a good idea, and a much better idea is to feed a tracking algorithm with the raw *sensor output* data directly. These new methods [24] require a precise modeling of the sensor physics and a direct access to the radar (or the sonar) raw data, i.e. to the signal intensity level in each azimuth/range cell. Note that these new methods seem well suited to new types of sensors such as lidar, since manufacturers do not integrate a detection module and do provide raw images of the signal intensity level in each azimuth/range cell.

The objective of the thesis is to study and design a tracking algorithm using raw data, and to implement it on radar (or sonar, or lidar) real data.

7. Partnerships and Cooperations

7.1. Regional initiatives

7.1.1. Stochastic Model-Data Coupled Representations for the Upper Ocean Dynamics (SEACS) — inter labex project

Participants: François Le Gland, Valérie Monbet.

January 2015 to December 2017.

This is a joint research initiative supported by the three labex active in Brittany, CominLabs (Communication and Information Sciences Laboratory), Lebesgue (Centre de Mathématiques Henri Lebesgue) and LabexMER (Frontiers in Marine Research).

This project aims at exploring novel statistical and stochastic methods to address the emulation, reconstruction and forecast of fine–scale upper ocean dynamics. The key objective is to investigate new tools and methods for the calibration and implementation of novel sound and efficient oceanic dynamical models, combining

- recent advances in the theoretical understanding, modeling and simulation of upper ocean dynamics,
- and mass of data routinely available to observe the ocean evolution.

In this respect, the emphasis will be given to stochastic frameworks to encompass multi–scale/multi–source approaches and benefit from the available observation and simulation massive data. The addressed scientific questions constitute basic research issues at the frontiers of several disciplines. It crosses in particular advanced data analysis approaches, physical oceanography and stochastic representations. To develop such an interdisciplinary initiative, the project gathers a set of research groups associated with these different scientific domains, which have already proven for several years their capacities to interact and collaborate on topics related to oceanic data and models. This project will place Brittany with an innovative and leading expertise at the frontiers of computer science, statistics and oceanography. This transdisciplinary research initiative is expected to resort to significant advances challenging the current thinking in computational oceanography.

7.2. National initiatives

7.2.1. Computational Statistics and Molecular Simulation (COSMOS) — ANR challenge Information and Communication Society

Participant: Frédéric Cérou.

Inria contract ALLOC 9452 — January 2015 to December 2017.

The COSMOS project 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. These two fields share some common history, but it seems that, in view of the quite recent specialization of the scientists and the techniques used in these respective fields, the communication between molecular simulation and computational statistics is not as intense as it should be.

We believe that there are therefore many opportunities in considering both fields at the same time: in particular, the adaption of a successful simulation technique from one field to the other requires first some abstraction process where the features specific to the original field of application are discarded and only the heart of the method is kept. Such a cross–fertilization is however only possible if the techniques developed in a specific field are sufficiently mature: this is why some fundamental studies specific to one of the application fields are still required. Our belief is that the embedding in a more general framework of specific developments in a given field will accelerate and facilitate the diffusion to the other field.

7.2.2. Advanced Geophysical Reduced–Order Model Construction from Image Observations (GERONIMO) — ANR programme Jeunes Chercheuses et Jeunes Chercheurs

Participant: Patrick Héas.

Inria contract ALLOC 8102 — March 2014 to February 2018.

The GERONIMO project aims at devising new efficient and effective techniques for the design 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. Our research activities are concerned by the exploitation of the huge amount of information contained in image data in order to reduce the uncertainty on the unknown parameters of the models and improve the reduced–model accuracy. In other words, the objective of our researches to process the large amount of incomplete and noisy image data daily captured by satellites sensors to devise new advanced model reduction techniques. The construction of ROMs is placed into a probabilistic Bayesian inference context, allowing for the handling of uncertainties associated to image measurements and the characterization of parameters of the reduced dynamical system.

7.3. European initiatives

7.3.1. Molecular Simulation: Modeling, Algorithms and Mathematical Analysis (MSMaths) — ERC Consolidator Grant

Participant: Mathias Rousset.

January 2014 to December 2019.

PI: Tony Lelièvre, Civil Engineer in Chief, Ecole des Ponts Paris-Tech.

Note that $1/3$ of Mathias Rousset research activities are held within the MSMath ERC project.

With the development of large–scale computing facilities, simulations of materials at the molecular scale are now performed on a daily basis. The aim of these simulations is to understand the macroscopic properties of matter from a microscopic description, for example, its atomistic configuration.

In order to make these simulations efficient and precise, mathematics have a crucial role to play. Indeed, specific algorithms have to be used in order to bridge the time and space scales between the atomistic level and the macroscopic level. The objective of the MSMath ERC project is thus to develop and study efficient algorithms to simulate high–dimensional systems over very long times. These developments are done in collaboration with physicists, chemists and biologists who are using these numerical methods in an academic or industrial context.

In particular, we are developping mathematical tools at the interface between the analysis of partial differential equations and stochastic analysis in order to characterize and to quantify the metastability of stochastic processes. Metastability is a fundamental concept to understand the timescale separation between the microscopic model and the macroscopic world. Many algorithms which aim at bridging the timescales are built using this timescale separation.

7.3.2. Design of Desalination Systems Based on Optimal Usage of Multiple Renewable Energy Sources (DESIRES) — ERANETMED NEXUS–14–049

Participant: Valérie Monbet.

January 2016 to December 2018.

This project is funded by the ERA–NET Initiative ERANETMED (Euro–Mediterranean Cooperation through ERA–NET Joint Activities and Beyond). It is a collaboration with Greece, Tunisia and Marocco, coordinated by Technical University of Crete (TUC). The French staff includes: Pierre Ailliot (Université de Bretagne Occidentale, Brest), Denis Allard (INRA Avignon), Anne Cuzol (Université de Bretagne Sud, Vannes), Christophe Maisondieu (IFREMER Brest) and Valérie Monbet.

The aim of DESIRES is to develop an Internet–based, multi–parametric electronic platform for optimum design of desalination plants, supplied by renewable energy sources (RES). The platform will rely upon (i) a solar, wind and wave energy potential database, (ii) existing statistical algorithms for processing energy-related data, (iii) information regarding the inter-annual water needs, (iv) a database with the technical characteristics of desalination plant units and the RES components, and (v) existing algorithms for cost effective design, optimal sizing and location selection of desalination plants.

7.4. International initiatives

7.4.1. Rare event simulation in epidemiology — PhD project at université de Ziguinchor

Participants: Ramatoulaye Dabo, François Le Gland.

This is the subjet of the PhD project of Ramatoulaye Dabo (université Assane Seck de Ziguinchor and université de Rennes 1).

The question here is to develop adaptive multilevel splitting algorithms for models that are commonly used in epidemiology, such as SIR (susceptible, infectious, recovered) models [26], or more complex compartmental models. A significant advantage of adaptive multilevel splitting is its robustness, since it does not require too much knowledge about the behavior of the system under study. An interesting challenge would be to understand how to couple the algorithm with numerically efficient simulation methods such as τ–leaping [35]. Complexity bounds and estimation error bounds could also be studied.

7.5. International research visitors

7.5.1. Visits to international teams

Patrick Héas has been invited to present his work on 3D wind field reconstruction by infrared sounding, at EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) in Darmstadt in February 2017.

8. Dissemination

8.1. Promoting scientific activities

8.1.1. Scientific events organisation

Valérie Monbet has co–organized the workshop and summer school on Data Science and Environment, held in Brest in July 2017. The conference gathered researchers that have an expertise in one of the two areas (data science, environmental data) and some interest for the other. Its main goal was to explore the fruitful interplay between the two areas, and ultimately to help create new connections and collaborations between the scientific communities involved. Another objective was to propose some high level courses and practices at the interaction of these two areas.

8.1.2. Participation in workshops, seminars, lectures, etc.

In addition to presentations with a publication in the proceedings, which are listed at the end of the document, members of ASPI have also given the following presentations.

Frédéric Cérou has given an invited talk on the convergence of adaptive multilevel splitting at the workshop Quasistationary Distributions: Analysis and Simulation held in Paderborn in September 2017.

Patrick Héas has presented his joint work with Mamadou Lamarana Diallo and Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique) on model reduction with “multi-space” prior information, at the European Conference on Numerical Mathematics and Advanced Applications (ENUMATH), held in Voss, Norway, in September 2017.

Thi Tuyet Trang Chau has presented her work on non parametric state–space model for missing–data imputation, at the workshop on Data Science and Environment, held in Brest in July 2017.

8.1.3. Research administration

François Le Gland is a member of the *conseil d'UFR* of the department of mathematics of université de Rennes 1. He is also a member of the *conseil scientifique* for the EDF/Inria scientific partnership.

Valérie Monbet is a member of both the *comité de direction* and the *conseil* of IRMAR (institut de recherche mathématiques de Rennes, UMR 6625). She is also the deputy head of the department of mathematics of université de Rennes 1, where she is a member of both the *conseil scientifique* and the *conseil d'UFR*.

8.2. Teaching, supervision, thesis committees

8.2.1. Teaching

Patrick Héas gives a course on Monte Carlo simulation methods in image analysis, at université de Rennes 1, within the SISEA (signal, image, systèmes embarqués, automatique) track of the master in electronical engineering and telecommunications.

François Le Gland gives

- 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, Brest.

Valérie Monbet gives

- a course on machine learning for biology at université de Rennes 1, within
  - the G2B (genetics, genomics, biochemistry) track of the master in molecular and cellular biology,
  - the MODE (modélisation en écologie) track of the master in biodiversity, ecology, evolution
  - and the master in scientific computing and modelling,
- a course on machine learning for environmental data, at the summer school on Data Science and Environment, held in Brest in July 2017,
- a course on graphical models at université de Rennes 1, within the master on applied mathematics and statistics,
- a course on MATLAB at université de Rennes 1, within the master in economics and financial engineering.

8.2.2. Supervision

François Le Gland and Valérie Monbet are jointly supervising one PhD student

- Thi Tuyet Trang Chau, provisional title: *Non parametric filtering for Metocean multi–source data fusion*, université de Rennes 1, started in October 2015, expected defense in October 2018, funding: Labex Lebesgue grant and Brittany council grant, co–direction: Pierre Ailliot (université de Bretagne Occidentale, Brest).

François Le Gland is supervising three other PhD students

- Kersane Zoubert–Ousseni, provisional title: *Particle filters for hybrid indoor navigation with smartphones*, université de Rennes 1, started in December 2014, expected defense in 2017, funding: CEA grant, co–direction: Christophe Villien (CEA LETI, Grenoble),
- Audrey Cuillery, provisional title: *Bayesian tracking from raw data*, université du Sud Toulon Var, started in April 2016, expected defense in 2019, funding: CIFRE grant with DCNS, co–direction: Claude Jauffret (université du Sud Toulon Var) and Dann Laneuville (DCNS, Nantes).
- Ramatoulaye Dabo, provisional title: *Rare event simulation in epidemiology*, université Assane Seck de Ziguinchor (Senegal) and université de Rennes 1, started in September 2015, expected defense in 2018, co–direction: Alassane Diedhiou (université Assane Seck de Ziguinchor).

Valérie Monbet is supervising two other PhD students

- Audrey Poterie, provisional title: *Régression d'une variable ordinale par des données longitudinales de grande dimension : application à la modélisation des effets secondaires suite à un traitement par radiothérapie*, université de Rennes 1, started in October 2015, expected defense in 2018, funding: INSA grant, co–direction: Jean–François Dupuy (INSA Rennes) and Laurent Rouvière (université de Haute Bretagne, Rennes).
- Marie Morvan, provisional title: *Modèles de régression pour données fonctionnelles. Application à la modélisation de données de spectrométrie dans le proche infra rouge*, université de Rennes 1, started in October 2016, expected defense in 2019, funding: MESR grant, co–direction: Joyce Giacofci (université de Haute Bretagne, Rennes) and Olivier Sire (université de Bretagne Sud, Vannes).

Mathias Rousset is supervising one PhD student

- Yushun Xu, provisional title: *Variance reduction of overdamped Langevin dynamics simulation*, université Paris-Est, started in October 2015, expected defense in 2018, co–direction: Pierre-André Zitt (université Paris–Est).

Patrick Héas has been supervising two post–doctoral fellows

- Hassan Maatouk, title: *Compressing the model by exploiting observations*, EPI ASPI, Inria Rennes–Bretagne Atlantique, started in September 2016, ended in September 2017, funding: ANR GERONIMO, co–supervision: Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique).
- Mamadou Lamarana Diallo, title: *Model reduction with "multi-space" prior information*, EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique, started in October 2016, ended in October 2017, funding: ANR GERONIMO, co–supervision: Cédric Herzet (EPI FLUMINANCE, Inria Rennes–Bretagne Atlantique).

8.2.3. Thesis committees

François Le Gland has been a member of the committee for the HDR of Christian Musso (université du Sud, Toulon).

Mathias Rousset has been a member of the committee for the PhD thesis of Gérôme Faure (CERMICS Ecole des Ponts Paris-Tech and CEA DAM, advisor: Gabriel Stoltz and Jean–Bernard Maillet).

9. Bibliography

Major publications by the team in recent years

[1] G. BIAU, F. CÉROU, A. GUYADER. *On the rate of convergence of the bagged nearest neighbor estimate*, in "Journal of Machine Learning Research", February 2010, vol. 11, pp. 687–712, http://jmlr.csail.mit.edu/papers/v11/biau10a.html

[2] G. BIAU, F. CÉROU, A. GUYADER. *On the rate of convergence of the functional k–nearest neighbor estimates*, in "IEEE Transactions on Information Theory", April 2010, vol. IT–56, n^{o} 4, pp. 2034–2040, http://dx.doi.org/10.1109/TIT.2010.2040857

[3] G. BIAU, F. CÉROU, A. GUYADER. *New insights into approximate Bayesian computation*, in "Annales de l'Institut Henri Poincaré, Probabilités et Statistiques", February 2015, vol. 51, n^{o} 1, pp. 376–403, http://dx.doi.org/10.1214/13-AIHP590

[4] F. CÉROU, A. GUYADER. *Nearest neighbor classification in infinite dimension*, in "ESAIM : Probability and Statistics", 2006, vol. 10, pp. 340–355, http://dx.doi.org/10.1051/ps:2006014

[5] F. CÉROU, A. GUYADER. *Adaptive multilevel splitting for rare event analysis*, in "Stochastic Analysis and Applications", March 2007, vol. 25, n^{o} 2, pp. 417–443, http://dx.doi.org/10.1080/07362990601139628

[6] F. CÉROU, P. DEL MORAL, F. LE GLAND, P. LEZAUD. *Genetic genealogical models in rare event analysis*, in "ALEA, Latin American Journal of Probability and Mathematical Statistics", 2006, vol. 1, pp. 181–203, Paper 01–08

[7] A. GUYADER, N. HENGARTNER, É. MATZNER–LØBER. *Simulation and estimation of extreme quantiles and extreme probabilities*, in "Applied Mathematics & Optimization", October 2011, vol. 64, n^{o} 2, pp. 171–196, http://dx.doi.org/10.1007/s00245-011-9135-z

[8] F. LE GLAND, V. MONBET, V.–D. TRAN. *Large sample asymptotics for the ensemble Kalman filter*, in "The Oxford Handbook of Nonlinear Filtering", Oxford, D. O. CRISAN, B. L. ROZOVSKII (editors), Oxford University Press, 2011, chap. 22, pp. 598–631

[9] F. LE GLAND, N. OUDJANE. *A sequential algorithm that keeps the particle system alive*, in "Stochastic Hybrid Systems : Theory and Safety Critical Applications", Berlin, H. A. P. BLOM, J. LYGEROS (editors), Lecture Notes in Control and Information Sciences, Springer–Verlag, 2006, n^{o} 337, pp. 351–389, http://dx.doi.org/10.1007/11587392_11

[10] C. MUSSO, N. OUDJANE, F. LE GLAND. *Improving regularized particle filters*, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 12, pp. 247–271, http://dx.doi.org/10.1007/978-1-4757-3437-9_12

Publications of the year

Articles in International Peer-Reviewed Journals

[11] 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

[12] 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

[13] M. LE CORVEC, C. JEZEQUEL, V. MONBET, N. FATIH, F. CHARPENTIER, H. TARIEL, C. BOUSSARD-PLÉDEL, B. BUREAU, O. LORÉAL, O. SIRE, E. BARDOU-JACQUET. *Mid-infrared spectroscopy of serum, a promising non-invasive method to assess prognosis in patients with ascites and cirrhosis*, in "PLoS ONE", October 2017, vol. 12, n^{o} 10, e0185997 p. [*DOI :* 10.1371/JOURNAL.PONE.0185997], https://hal.archives-ouvertes.fr/hal-01625051

[14] V. MONBET, P. AILLIOT. *Sparse vector Markov switching autoregressive models. Application to multivariate time series of temperature*, in "Computational Statistics and Data Analysis", 2017, vol. 108, pp. 40-51 [*DOI :* 10.1016/J.CSDA.2016.10.023], https://hal.archives-ouvertes.fr/hal-01250058

International Conferences with Proceedings

[15] P. HÉAS, C. HERZET. *Optimal Low-Rank Dynamic Mode Decomposition*, in "Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans 2017", New Orleans, United States, March 2017, https://arxiv.org/abs/1701.01064 [*DOI :* 10.1109/ICASSP.2017.7952999], https://hal.archives-ouvertes.fr/hal-01429975

[16] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. *Seismic induced damage detection through parallel estimation of force and parameter using improved interacting Particle-Kalman filter*, in "11th International Workshop on Structural Health Monitoring", San Francisco, United States, September 2017, https://hal.inria.fr/hal-01590713

[17] K. ZOUBERT-OUSSENI, C. VILLIEN, F. LE GLAND. *Post-processing optimization of piecewise indoor trajectories based on IMU and RSS measurements*, in "Proceedings of the 8th International Conference on Indoor Positioning and Indoor Navigation (IPIN)", Sapporo, Japan, IEEE, September 2017 [*DOI :* 10.1109/IPIN.2017.8115903], https://hal.inria.fr/hal-01671149

Conferences without Proceedings

[18] S. SEN, A. CRINIÈRE, L. MEVEL, F. CÉROU, J. DUMOULIN. *Estimation of time varying system parameters from ambient response using improved Particle-Kalman filter with correlated noise*, in "EGU General Assembly 2017", Vienne, Austria, April 2017, https://hal.inria.fr/hal-01493840

Other Publications

[19] B. DELYON, F. CÉROU, A. GUYADER, M. ROUSSET. *A Central Limit Theorem for Fleming-Viot Particle Systems with Hard Killing*, October 2017, https://arxiv.org/abs/1709.06771 - 38 pages, https://hal.archives-ouvertes.fr/hal-01614094

[20] P. HÉAS, C. HERZET. *Optimal Kernel-Based Dynamic Mode Decomposition*, 2017, https://arxiv.org/abs/1710.10919 - working paper or preprint [*DOI :* 10.10919], https://hal.archives-ouvertes.fr/hal-01634542

[21] M. ROUSSET, Y. XU, P.-A. ZITT. *A Weak Overdamped Limit Theorem for Langevin Processes*, September 2017, https://arxiv.org/abs/1709.09866 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01596954

References in notes

[22] A. DOUCET, N. DE FREITAS, N. J. GORDON (editors). *Sequential Monte Carlo methods in practice*, Statistics for Engineering and Information Science, Springer–Verlag, New York, 2001, http://dx.doi.org/10.1007/978-1-4757-3437-9

[23] M. S. ARULAMPALAM, S. MAKSELL, N. J. GORDON, T. CLAPP. *A tutorial on particle filters for online nonlinear / non–Gaussian Bayesian tracking*, in "IEEE Transactions on Signal Processing", February 2002, vol. SP–50, n° 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), pp. 174–188, http://dx.doi.org/10.1109/78.978374

[24] M. BOCQUEL. *Random finite sets in multi–target tracking: efficient sequential MCMC implementation*, Department of Applied Mathematics, University of Twente, Enschede, October 2013, http://purl.utwente.nl/publications/87496

[25] O. CAPPÉ, S. J. GODSILL, É. MOULINES. *An overview of existing methods and recent advances in sequential Monte Carlo*, in "Proceedings of the IEEE", May 2007, vol. 95, n° 5 (Special issue on Large–Scale Dynamic Systems), pp. 899–924, http://dx.doi.org/10.1109/JPROC.2007.893250

[26] D. J. DALEY, J. GANI. *Epidemic modelling: An introduction*, Cambridge Studies in Mathematical Biology, Cambridge University Press, Cambridge, 1999, vol. 15, http://dx.doi.org/10.1017/CBO9780511608834

[27] P.–T. DE BOER, D. P. KROESE, S. MANNOR, R. Y. RUBINSTEIN. *A tutorial on the cross–entropy method*, in "Annals of Operations Research", January 2005, vol. 134 (Special issue on the Cross-Entropy Method for Combinatorial Optimization, Rare Event Simulation and Neural Computation), n° 1, pp. 19–67, http://dx.doi.org/10.1007/s10479-005-5724-z

[28] P. DEL MORAL. *Mean field simulation for Monte Carlo integration*, Monographs on Statistics and Applied Probability, Chapman & Hall / CRC Press, London, 2013, vol. 126, http://www.crcpress.com/product/isbn/9781466504059

[29] P. DEL MORAL. *Feynman–Kac formulae. Genealogical and interacting particle systems with applications*, Probability and its Applications, Springer–Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4684-9393-1

[30] P. DEL MORAL, L. MICLO. *Branching and interacting particle systems approximations of Feynman–Kac formulae with applications to nonlinear filtering*, in "Séminaire de Probabilités XXXIV", Berlin, J. AZÉMA, M. ÉMERY, M. LEDOUX, M. YOR (editors), Lecture Notes in Mathematics, Springer–Verlag, 2000, vol. 1729, pp. 1–145, http://dx.doi.org/10.1007/BFb0103798

[31] R. DOUC, É. MOULINES. *Limit theorems for weighted samples with applications to sequential Monte Carlo methods*, in "The Annals of Statistics", October 2008, vol. 36, n° 5, pp. 2344–2376, http://dx.doi.org/10.1214/07-AOS514

[32] D. FOX, J. HIGHTOWER, L. LIAO, D. SCHULZ, G. BORRIELLO. *Bayesian filtering for location estimation*, in "IEEE Pervasive Computing", July/September 2003, vol. 2, n^{O} 3, pp. 24–33, http://dx.doi.org/10.1109/MPRV.2003.1228524

[33] D. FOX, S. THRUN, W. BURGARD, F. DELLAERT. *Particle filters for mobile robot localization*, in "Sequential Monte Carlo Methods in Practice", New York, A. DOUCET, N. DE FREITAS, N. J. GORDON (editors), Statistics for Engineering and Information Science, Springer–Verlag, 2001, chap. 19, pp. 401–428, http://dx.doi.org/10.1007/978-1-4757-3437-9_19

[34] D. FRENKEL, B. SMIT. *Understanding molecular simulation. From algorithms to applications*, Computational Science Series, 2nd, Academic Press, San Diego, 2002, vol. 1, http://www.sciencedirect.com/science/book/9780122673511

[35] D. T. GILLESPIE. *Approximate accelerated stochastic simulation of chemically reacting systems*, in "The Journal of Chemical Physics", July 2001, vol. 115, n^{O} 4, pp. 1716–1733, http://dx.doi.org/10.1063/1.1378322

[36] P. GLASSERMAN. *Monte Carlo methods in financial engineering*, Applications of Mathematics, Springer–Verlag, New York, 2004, vol. 53, http://dx.doi.org/10.1007/978-0-387-21617-1

[37] P. GLASSERMAN, P. HEIDELBERGER, P. SHAHABUDDIN, T. ZAJIC. *Multilevel splitting for estimating rare event probabilities*, in "Operations Research", July–August 1999, vol. 47, n^{O} 4, pp. 585–600, http://dx.doi.org/10.1287/opre.47.4.585

[38] N. J. GORDON, D. J. SALMOND, A. F. M. SMITH. *Novel approach to nonlinear / non–Gaussian Bayesian state estimation*, in "IEE Proceedings, Part F", April 1993, vol. 140, n^{O} 2, pp. 107–113, http://dx.doi.org/10.1049/ip-f-2.1993.0015

[39] F. GUSTAFSSON, F. GUNNARSSON, N. BERGMAN, U. FORSSELL, J. JANSSON, R. KARLSSON, P.–J. NORDLUND. *Particle filters for positioning, navigation, and tracking*, in "IEEE Transactions on Signal Processing", February 2002, vol. SP–50, n^{O} 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), pp. 425–437, http://dx.doi.org/10.1109/78.978396

[40] J. HOUSSINEAU. *Representation and estimation of stochastic populations*, School of Engineering and Physical Sciences, Heriot–Watt University, Edinburgh, August 2015, https://jeremiehoussineau.wordpress.com/ph-d-thesis-all-versions/

[41] P. HÉAS, C. HERZET. *Low–Rank Dynamic Mode Decomposition: Optimal Solution in Polynomial–Time*, in "ArXiv e-prints", October 2016, https://arxiv.org/abs/1610.02962

[42] M. ISARD, A. BLAKE. CONDENSATION — *Conditional density propagation for visual tracking*, in "International Journal of Computer Vision", August 1998, vol. 29, n^{O} 1, pp. 5–28, http://dx.doi.org/10.1023/A:1008078328650

[43] G. KITAGAWA. *Monte Carlo filter and smoother for non–Gaussian nonlinear state space models*, in "Journal of Computational and Graphical Statistics", 1996, vol. 5, n^{O} 1, pp. 1–25, http://www.jstor.org/stable/1390750

[44] H. R. KÜNSCH. *Recursive Monte Carlo filters : Algorithms and theoretical analysis*, in "The Annals of Statistics", October 2005, vol. 33, n^{O} 5, pp. 1983–2021, http://www.jstor.org/stable/3448632

[45] P. L'ÉCUYER, V. DEMERS, B. TUFFIN. *Rare events, splitting, and quasi–Monte Carlo*, in "ACM Transactions on Modeling and Computer Simulation", April 2007, vol. 17, n° 2 (Special issue honoring Perwez Shahabuddin), Article 9, http://dx.doi.org/10.1145/1225275.1225280

[46] F. LE GLAND, N. OUDJANE. *A robustification approach to stability and to uniform particle approximation of nonlinear filters : the example of pseudo-mixing signals*, in "Stochastic Processes and their Applications", August 2003, vol. 106, n° 2, pp. 279-316, http://dx.doi.org/10.1016/S0304-4149(03)00041-3

[47] F. LE GLAND, N. OUDJANE. *Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters*, in "The Annals of Applied Probability", February 2004, vol. 14, n° 1, pp. 144–187, http://dx.doi.org/10.1214/aoap/1075828050

[48] J. S. LIU. *Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer–Verlag, New York, 2001, http://dx.doi.org/10.1007/978-0-387-76371-2

[49] H. LOUVIN, E. DUMONTEIL, T. LELIÈVRE, M. ROUSSET, C. M. DIOP. *Adaptive multilevel splitting for Monte Carlo particle transport*, in "EPJ Web of Conferences", 2017, vol. 153 [*DOI :* 10.1051/EPJCONF/201715306006], https://hal.archives-ouvertes.fr/hal-01661012

[50] R. P. S. MAHLER. *Multitarget Bayes filtering via first–order multitarget moments*, in "IEEE Transactions on Aerospace and Electronic Systems", October 2003, vol. AES–39, n° 4, pp. 1152–1178, http://dx.doi.org/10.1109/TAES.2003.1261119

[51] B. RISTIĆ, M. S. ARULAMPALAM, N. J. GORDON. *Beyond the Kalman Filter : Particle Filters for Tracking Applications*, Artech House, Norwood, MA, 2004, http://www.artechhouse.com/international/books/945.aspx

[52] R. Y. RUBINSTEIN, D. P. KROESE. *The cross–entropy method. A unified approach to combinatorial optimization, Monte Carlo simulation and machine learning*, Information Science and Statistics, Springer–Verlag, New York, 2004, http://dx.doi.org/10.1007/978-1-4757-4321-0

[53] C. J. STONE. *Consistent nonparametric regression (with discussion)*, in "The Annals of Statistics", July 1977, vol. 5, n° 4, pp. 595–645, http://www.jstor.org/stable/2958783

[54] L. D. STONE, R. L. STREIT, T. L. CORWIN, K. L. BELL. *Bayesian multiple target tracking*, 2nd, Artech House, Norwood, MA, 2014, http://www.artechhouse.com/international/books/2090.aspx

[55] S. THRUN, W. BURGARD, D. FOX. *Probabilistic robotics*, Intelligent Robotics and Autonomous Agents, The MIT Press, Cambridge, MA, 2005, http://mitpress.mit.edu/books/probabilistic-robotics

[56] B.–N. VO, W.–K. MA. *The Gaussian mixture probability hypothesis density filter*, in "IEEE Transactions on Signal Processing", November 2006, vol. SP–54, n° 11, pp. 4091–4104, http://dx.doi.org/10.1109/TSP.2006.881190

[57] K. ZOUBERT-OUSSENI, C. VILLIEN, F. LE GLAND. *Comparison of post-processing algorithms for indoor navigation trajectories*, in "7th International Conference on Indoor Positioning and Indoor Navigation (IPIN)", Alcala de Henares, Spain, IEEE, October 2016 [*DOI :* 10.1109/IPIN.2016.7743590], https://hal.inria.fr/hal-01423198

[58] A. W. VAN DER VAART, J. A. WELLNER. *Weak convergence and empirical processes*, Springer Series in Statistics, Springer–Verlag, Berlin, 1996, http://dx.doi.org/10.1007/978-1-4757-2545-2