



IN PARTNERSHIP WITH:  
**CNRS**

**Université Haute Bretagne  
(Rennes 2)**

**Université Rennes 1**

Activity Report 2015

## **Project-Team ASPI**

Applications of interacting particle systems to  
statistics

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

RESEARCH CENTER  
**Rennes - Bretagne-Atlantique**

THEME  
**Stochastic approaches**



## Table of contents

<b>1. Members</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Research Program</b>	<b>2</b>
3.1. Interacting Monte Carlo methods and particle approximation of Feynman–Kac distributions	2
3.2. Statistics of HMM	4
3.3. Multilevel splitting for rare event simulation	5
3.4. Statistical learning: pattern recognition and nonparametric regression	7
<b>4. Application Domains</b>	<b>8</b>
4.1. Localisation, navigation and tracking	8
4.2. Rare event simulation	9
<b>5. New Results</b>	<b>9</b>
5.1. Adaptive multilevel splitting	9
5.2. Adaptive multilevel splitting as a Fleming–Viot system	9
5.3. Bias and variance reduction in rare event simulation	9
5.4. Simulation–based algorithms for the optimization of sensor deployment	10
5.5. Kalman Laplace filtering	10
5.6. Combining analog method and ensemble data assimilation	10
5.7. Markov–switching vector autoregressive models	11
5.8. Dependent time changed processes	11
5.9. An efficient algorithm for video super–resolution based on a sequential model	12
5.10. Reduced–order modeling of hidden dynamics	12
<b>6. Bilateral Contracts and Grants with Industry</b>	<b>12</b>
6.1. Bilateral contracts with industry	12
6.2. Bilateral grants with industry	13
<b>7. Partnerships and Cooperations</b>	<b>13</b>
7.1. Regional initiatives	13
7.2. National initiatives	14
7.2.1. Computational Statistics and Molecular Simulation (COSMOS) — ANR challenge Information and Communication Society	14
7.2.2. Advanced Geophysical Reduced–Order Model Construction from Image Observations (GERONIMO) — ANR programme Jeunes Chercheuses et Jeunes Chercheurs	14
7.3. International research visitors	14
<b>8. Dissemination</b>	<b>14</b>
8.1. Promoting scientific activities	14
8.1.1. Scientific events organisation	14
8.1.2. Journal	15
8.1.3. Invited talks	15
8.2. Teaching, supervision, thesis committees	15
8.2.1. Teaching	15
8.2.2. Supervision	15
8.2.3. Thesis committees	16
8.3. Participation in workshops, seminars, lectures, etc.	16
<b>9. Bibliography</b>	<b>16</b>



## Project-Team ASPI

*Creation of the Project-Team: 2005 January 10*

### Keywords:

#### **Computer Science and Digital Science:**

- 3.4.4. - Optimization and learning
- 3.4.5. - Bayesian methods
- 5.4.5. - Object tracking and motion analysis
- 5.4.6. - Object localization
- 5.9.2. - Estimation, modeling
- 6.1.2. - Stochastic Modeling (SPDE, SDE)
- 6.2.2. - Numerical probability
- 6.2.3. - Probabilistic methods
- 6.2.4. - Statistical methods
- 6.3.2. - Data assimilation
- 6.3.4. - Model reduction

#### **Other Research Topics and Application Domains:**

- 3.2. - Climate and meteorology
- 3.3.2. - Water: sea & ocean, lake & river
- 7.1.3. - Air traffic
- 9.4.4. - Chemistry
- 9.9. - Risk management

## 1. Members

### **Research Scientists**

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

### **Faculty Member**

Valérie Monbet [Univ. Rennes I, Professor, HdR]

### **PhD Students**

Kersane Zoubert-Ousseni [CEA]  
Chau Thi Tuyet Trang [Univ. Rennes 1, since Oct 2015]  
Audrey Poterie [Univ. Rennes 1, since Oct 2015]

### **Visiting Scientist**

Arnaud Guyader [Univ. Paris 6, Professor, HdR]

### **Administrative Assistant**

Fabienne Cuyollaa [Inria]

## 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) [46], 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 recent and major breakthrough, 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* [49], *Monte Carlo filter* [54] or *condensation* (conditional density propagation) algorithm [51], and which historically has been the first algorithm to include a redistribution 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 [50], [43], visual tracking [51], mobile robotics [44], [66], ubiquitous computing and ambient intelligence, sensor networks, risk evaluation and simulation of rare events [47], genetics, molecular simulation [45], etc.

Other examples of the many applications of particle filtering can be found in the contributed volume [30] 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 [31] can be found, and in the textbook [63] 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 [62]. A recent overview can also be found in [32].

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 [38], [36], 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 [68]), see e.g. the survey article [38] or the textbooks [36], [35], and references therein

convergence in  $\mathbb{L}^p$ , convergence as empirical processes indexed by classes of functions, uniform convergence in time, see also [59], [60], central limit theorem, see also [56], [40], 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. Statistics of HMM

Hidden Markov models (HMM) form a special case of partially observed stochastic dynamical systems, in which the state of a Markov process (in discrete or continuous time, with finite or continuous state space) should be estimated from noisy observations. The conditional probability distribution of the hidden state given past observations is a well-known example of a normalized (nonlinear) Feynman–Kac distribution, see 3.1. These models are very flexible, because of the introduction of latent variables (non observed) which allows to model complex time dependent structures, to take constraints into account, etc. In addition, the underlying Markovian structure makes it possible to use numerical algorithms (particle filtering, Markov chain Monte Carlo methods (MCMC), etc.) which are computationally intensive but whose complexity is rather small. Hidden Markov models are widely used in various applied areas, such as speech recognition, alignment of biological sequences, tracking in complex environment, modeling and control of networks, digital communications, etc.

Beyond the recursive estimation of a hidden state from noisy observations, the problem arises of statistical inference of HMM with general state space [33], [41], including estimation of model parameters, early monitoring and diagnosis of small changes in model parameters, etc.

**Large time asymptotics** A fruitful approach is the asymptotic study, when the observation time increases to infinity, of an extended Markov chain, whose state includes (i) the hidden state, (ii) the observation, (iii) the prediction filter (i.e. the conditional probability distribution of the hidden state given observations at all previous time instants), and possibly (iv) the derivative of the prediction filter with respect to the parameter. Indeed, it is easy to express the log-likelihood function, the conditional least-squares criterion, and many other classical contrast processes, as well as their derivatives with respect to the parameter, as additive functionals of the extended Markov chain.

The following general approach has been proposed

- first, prove an exponential stability property (i.e. an exponential forgetting property of the initial condition) of the prediction filter and its derivative, for a misspecified model,



- from this, deduce a geometric ergodicity property and the existence of a unique invariant probability distribution for the extended Markov chain, hence a law of large numbers and a central limit theorem for a large class of contrast processes and their derivatives, and a local asymptotic normality property,
- finally, obtain the consistency (i.e. the convergence to the set of minima of the associated contrast function), and the asymptotic normality of a large class of minimum contrast estimators.

This programme has been completed in the case of a finite state space [8], and has been generalized [39] under an uniform minoration assumption for the Markov transition kernel, which typically does only hold when the state space is compact. Clearly, the whole approach relies on the existence of an exponential stability property of the prediction filter, and the main challenge currently is to get rid of this uniform minoration assumption for the Markov transition kernel [37], [60], so as to be able to consider more interesting situations, where the state space is noncompact.

**Small noise asymptotics** Another asymptotic approach can also be used, where it is rather easy to obtain interesting explicit results, in terms close to the language of nonlinear deterministic control theory [55]. Taking the simple example where the hidden state is the solution to an ordinary differential equation, or a nonlinear state model, and where the observations are subject to additive Gaussian white noise, this approach consists in assuming that covariances matrices of the state noise and of the observation noise go simultaneously to zero. If it is reasonable in many applications to consider that noise covariances are small, this asymptotic approach is less natural than the large time asymptotics, where it is enough (provided a suitable ergodicity assumption holds) to accumulate observations and to see the expected limit laws (law of large numbers, central limit theorem, etc.). In opposition, the expressions obtained in the limit (Kullback–Leibler divergence, Fisher information matrix, asymptotic covariance matrix, etc.) take here a much more explicit form than in the large time asymptotics.

The following results have been obtained using this approach

- the consistency of the maximum likelihood estimator (i.e. the convergence to the set  $M$  of global minima of the Kullback–Leibler divergence), has been obtained using large deviations techniques, with an analytical approach [52],
- if the abovementioned set  $M$  does not reduce to the true parameter value, i.e. if the model is not identifiable, it is still possible to describe precisely the asymptotic behavior of the estimators [53]: in the simple case where the state equation is a noise-free ordinary differential equation and using a Bayesian framework, it has been shown that (i) if the rank  $r$  of the Fisher information matrix  $I$  is constant in a neighborhood of the set  $M$ , then this set is a differentiable submanifold of codimension  $r$ , (ii) the posterior probability distribution of the parameter converges to a random probability distribution in the limit, supported by the manifold  $M$ , absolutely continuous w.r.t. the Lebesgue measure on  $M$ , with an explicit expression for the density, and (iii) the posterior probability distribution of the suitably normalized difference between the parameter and its projection on the manifold  $M$ , converges to a mixture of Gaussian probability distributions on the normal spaces to the manifold  $M$ , which generalized the usual asymptotic normality property,
- it has been shown [61] that (i) the parameter dependent probability distributions of the observations are locally asymptotically normal (LAN) [58], from which the asymptotic normality of the maximum likelihood estimator follows, with an explicit expression for the asymptotic covariance matrix, i.e. for the Fisher information matrix  $I$ , in terms of the Kalman filter associated with the linear tangent linear Gaussian model, and (ii) the score function (i.e. the derivative of the log-likelihood function w.r.t. the parameter), evaluated at the true value of the parameter and suitably normalized, converges to a Gaussian r.v. with zero mean and covariance matrix  $I$ .

### 3.3. Multilevel splitting for rare event simulation

See 4.2, and 5.1, 5.2, 5.3, and 5.4.

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 inefficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [57], 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 [5], 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 method 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 using an artificial Markovian dynamics. 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, conditioned 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 redistribution. 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 [64], [34], 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.

### 3.4. 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 [65] 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 [3] 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, ENS Paris and EPI CLASSIC, Inria Paris—Rocquencourt). 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

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 [50], [43], 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 [49], [63], 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 3.3, and 5.1, 5.2, 5.3, and 5.4.

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 un-efficient to estimate accurately very small probabilities. Besides importance sampling, an alternate widespread technique consists in multilevel splitting [57], 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. Adaptive multilevel splitting

**Participants:** Frédéric Cérou, Arnaud Guyader.

We have show last year that an adaptive version of multilevel splitting for rare events is strongly consistent and that the estimates satisfy a CLT (central limit theorem), with the same asymptotic variance as the non-adaptive algorithm with the optimal choice of the parameters. This year we have generalized these results to include Markov kernels used to move the particles (or *shakers*) are of Metropolis–Hastings type. This is a non-trivial generalization to a very important case.

### 5.2. Adaptive multilevel splitting as a Fleming–Viot system

**Participants:** Frédéric Cérou, Arnaud Guyader.

This is a collaboration with Bernard Delyon (université de Rennes 1) and Mathias Rousset (EPI MATHERIALS, Inria Paris Rocquencourt).

By considering the adaptive multilevel splitting algorithm as a Fleming–Viot particle system for a stochastic wave, in the sense of [42], we have shown the mean square convergence using a general result [67] about the convergence of Fleming–Viot (Villemonais, 2013). We are currently working on the proof of a central limit theorem, but the proof is not yet complete. We have nevertheless identified the expression of the asymptotic variance.

### 5.3. Bias and variance reduction in rare event simulation

**Participant:** François Le Gland.

This is a collaboration with Damien Jacquemart (ONERA, Palaiseau) and Jérôme Morio (ONERA, Toulouse).

In [17], we highlight a bias induced by the discretization of the sampled Markov paths in the splitting algorithm, and we propose to correct this bias using a deformation of the intermediate regions, as proposed in [48]. Moreover, we propose two numerical methods to design intermediate regions in the splitting algorithm that minimise the variance. One is connected with a partial differential equation approach, the other one is based on the discretization of the state space of the process.

## 5.4. Simulation-based algorithms for the optimization of sensor deployment

**Participant:** François Le Gland.

This is a collaboration with Christian Musso (ONERA, Palaiseau) and with Sébastien Paris (LSIS, université du Sud Toulon Var).

The problem considered here can be described as follows: a limited number of sensors should be deployed by a carrier in a given area, and should be activated at a limited number of time instants within a given time period, so as to maximize the probability of detecting a target (present in the given area during the given time period). There is an information dissymmetry in the problem: if the target is sufficiently close to a sensor position when it is activated, then the target can learn about the presence and exact position of the sensor, and can temporarily modify its trajectory so as to escape away before it is detected. This is referred to as the target intelligence. Two different simulation-based algorithms have been designed in [23] to solve separately or jointly this optimization problem, with different and complementary features. One is fast, and sequential: it proceeds by running a population of targets and by dropping and activating a new sensor (or re-activating a sensor already available) where and when this action seems appropriate. The other is slow, iterative, and non-sequential: it proceeds by updating a population of deployment plans with guaranteed and increasing criterion value at each iteration, and for each given deployment plan, there is a population of targets running to evaluate the criterion. Finally, the two algorithms can cooperate in many different ways, to try and get the best of both approaches. A simple and efficient way is to use the deployment plans provided by the sequential algorithm as the initial population for the iterative algorithm.

## 5.5. Kalman Laplace filtering

**Participant:** François Le Gland.

This is a collaboration with Paul Bui Quang (CEA, Bruyères-le-Châtel) and Christian Musso (ONERA, Palaiseau).

We propose in [21] a new nonlinear Bayesian filtering algorithm where the prediction step is performed like in the extended Kalman filter, and the update step is done thanks to the Laplace method for integral approximation. This algorithm is called the Kalman Laplace filter (KLF). The KLF provides a closed-form non-Gaussian approximation of the posterior density. The hidden state is estimated by the maximum a posteriori. We describe a way to alleviate the computation cost of this maximization, when the likelihood is a function of a vector whose dimension is smaller than the state space dimension. The KLF is tested on three simulated nonlinear filtering problems: target tracking with angle measurements, population dynamics monitoring, motion reconstruction by neural decoding. It exhibits a good performance, especially when the observation noise is small.

## 5.6. Combining analog method and ensemble data assimilation

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

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

Nowadays, ocean and atmosphere sciences face a deluge of data from spatial observations, in situ monitoring as well as numerical simulations. The availability of these different data sources offer new opportunities, still largely underexploited, to improve the understanding, modeling and reconstruction of geophysical dynamics. The classical way to reconstruct the space-time variations of a geophysical system from observations relies on data assimilation methods using multiple runs of the known dynamical model. This classical framework may have severe limitations including its computational cost, the lack of adequacy of the model with observed data, modeling uncertainties. In [24], we explore an alternative approach and develop a fully data-driven framework, which combines machine learning and statistical sampling to simulate the dynamics of complex system. As a proof concept, we address the assimilation of the chaotic Lorenz-63 model. We demonstrate that



a nonparametric sampler from a catalog of historical datasets, namely a nearest neighbor or analog sampler, combined with a classical stochastic data assimilation scheme, the ensemble Kalman filter and smoother, reach state-of-the-art performances, without online evaluations of the physical model.

## 5.7. Markov-switching vector autoregressive models

**Participant:** Valérie Monbet.

This is a collaboration with Pierre Ailliot (université de Bretagne Occidentale), Julie Bessac (Argonne National Laboratory, Chicago) and Julien Cattiaux (Météo-France, Toulouse).

Multivariate time series are of interest in many fields including economics and environment. The most popular tools for studying multivariate time series are the vector autoregressive (VAR) models because of their simple specification and the existence of efficient methods to fit these models. However, the VAR models do not allow to describe time series mixing different dynamics. For instance, when meteorological variables are observed, the resulting time series exhibit an alternance of different temporal dynamics corresponding to weather regimes. The regime is often not observed directly and is thus introduced as a latent process in time series models in the spirit of hidden Markov models. Markov switching vector autoregressive (MSVAR) models have been introduced as a generalization of autoregressive models and hidden Markov models. They lead to flexible and interpretable models. In this multivariate context, several questions occur.

- The discrete hidden variable also called regime has to be correctly defined. Indeed the regime can be local (e.g. link to a subset of the variables) or global (e.g. the same for all the variables). It can also be observed and inferred a priori or hidden. In the second case, it has to be estimated at the same time as the model parameters.

The question of the definition of the regime is investigated in [26] for the specific problem of multi site wind modeling.

- Markov Switching VAR models (MSVAR) suffer of the same dimensionality problem as VAR models. For large (and even moderate) dimensions, the number of autoregressive coefficients in each regime can be prohibitively large which results in noisy estimates. When the variables are correlated, which is the standard situation in multivariate time series, over-learning is frequent. The estimated parameters contains spurious non-zero coefficients and are then difficult to interpret. The predictions associated to the model are usually unstable. Collinearity causes also ill-conditioning of the innovation covariance. In [29], we propose a likelihood penalization method with hard thresholding for MSVAR models leading to sparse MSVAR. Both autoregressive matrices and precision matrices are penalized using smoothly clipped absolute deviation (SCAD) penalties.

## 5.8. Dependent time changed processes

**Participant:** Valérie Monbet.

This is a collaboration with Pierre Ailliot (université de Bretagne Occidentale), Bernard Delyon (université de Rennes 1) and Marc Prevosto (IFREMER, Brest).

Many records in environmental sciences exhibit asymmetric trajectories and there is a need for simple and tractable models which can reproduce such feature. In [25] we explore an approach based on applying both a time change and a marginal transformation on Gaussian processes. The main originality of the proposed model is that the time change depends on the observed trajectory. We first show that the proposed model is stationary and ergodic and provide an explicit characterization of the stationary distribution. This result is then used to build both parametric and non-parametric estimate of the time change function whereas the estimation of the marginal transformation is based on up-crossings. Simulation results are provided to assess the quality of the estimates. The model is applied to wave data and it is shown that the fitted model is able to reproduce important statistics of the data such as its spectrum and marginal distribution which are important quantities for practical applications. An important benefit of the proposed model is its ability to reproduce the observed asymmetries between the crest and the troughs and between the front and the back of the waves by accelerating the chronometer in the crests and in the front of the waves.

## 5.9. An efficient algorithm for video super-resolution based on a sequential model

**Participant:** Patrick Héas.

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

In the work [27], we propose a novel procedure for video super-resolution, that is the recovery of a sequence of high-resolution images from its low-resolution counterpart. Our approach is based on a *sequential* model (i.e. each high-resolution frame is supposed to be a displaced version of the preceding one) and considers the use of sparsity-enforcing priors. Both the recovery of the high-resolution images and the motion fields relating them is tackled. This leads to a large-dimensional, non-convex and non-smooth problem. We propose an algorithmic framework to address the latter. Our approach relies on fast gradient evaluation methods and modern optimization techniques for non-differentiable/non-convex problems. Unlike some other previous works, we show that there exists a provably-convergent method with a complexity linear in the problem dimensions. We assess the proposed optimization method on several video benchmarks and emphasize its good performance with respect to the state of the art.

## 5.10. Reduced-order modeling of hidden dynamics

**Participant:** Patrick Héas.

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

The objective of the paper [28] is to investigate how noisy and incomplete observations can be integrated in the process of building a reduced-order model. This problematic arises in many scientific domains where there exists a need for accurate low-order descriptions of highly-complex phenomena, which can not be directly and/or deterministically observed. Within this context, the paper proposes a probabilistic framework for the construction of POD–Galerkin reduced-order models. Assuming a hidden Markov chain, the inference integrates the uncertainty of the hidden states relying on their posterior distribution. Simulations show the benefits obtained by exploiting the proposed framework.

# 6. Bilateral Contracts and Grants with Industry

## 6.1. Bilateral contracts with industry

### 6.1.1. Optimization of sensors location and activation (DUCATI) — contract with DGA / Techniques navales

**Participant:** François Le Gland.

See 3.3, 4.2 and 5.4

*Inria contract ALLOC 7326 — April 2013 to December 2016.*

This is a collaboration with Christian Musso (ONERA, Palaiseau) and with Sébastien Paris (LSIS, université du Sud Toulon Var).

The objective of this project is to optimize the position and activation times of a few sensors deployed by one or several platforms over a search zone, so as to maximize the probability of detecting a moving target. The difficulty here is that the target can detect an activated sensor before it is detected itself, and it can then modify its own trajectory to escape from the sensor. This makes the optimization problem a spatio-temporal problem. Our contribution has been to study different ways to merge two different solutions to the optimization problem : a fast, though suboptimal, solution developed by ONERA in which sensors are deployed where and when the probability of presence of a target is high enough, and the optimal population-based solution developed by LSIS and Inria in a previous contract (Inria contract ALLOC 4233) with DGA / Techniques navales.



## 6.2. Bilateral grants with industry

### 6.2.1. Hybrid indoor navigation — PhD grant 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.

## 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 is 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. International research visitors

### 7.3.1. *Visits to international teams*

François Le Gland has been invited by Joaquín Míguez to visit the department of signal theory and communications of Universidad Carlos III de Madrid, in February 2015.

## 8. Dissemination

### 8.1. Promoting scientific activities

#### 8.1.1. *Scientific events organisation*

Valérie Monbet has co-organized the workshop on *Stochastic Model-Data Coupled Representations for the Upper Ocean Dynamics*, the kick-off meeting of the SEACS project, held in Landeda in May 2015.

### 8.1.2. Journal

Valérie Monbet has been the guest editor of a special issue (volume 156, number 1) on stochastic weather generators, in *Journal de la Société Française de Statistique*.

### 8.1.3. Invited talks

Valérie Monbet has given an invited talk on Markov-switching vector autoregressive models for multivariate time series of air temperature, at *47èmes Journées de Statistique*, held in Lille in June 2015.

## 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, école doctorale MATISSE) track of the master in electrical engineering and telecommunications.

François Le Gland gives

- 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 electrical engineering and telecommunications,
- a 3rd year course on **Bayesian filtering and particle approximation**, at ENSTA (école nationale supérieure de techniques avancées), Paris, within the systems and control module,
- a 3rd year course on **linear and nonlinear filtering**, at ENSAI (école nationale de la statistique et de l'analyse de l'information), Ker Lann, within the statistical engineering track,
- and a 3rd year course on **hidden Markov models**, at Télécom Bretagne, Brest.

Valérie Monbet gives several courses on data analysis, on time series, and on mathematical statistics, all at université de Rennes 1 within the master on statistics and econometrics.

### 8.2.2. Supervision

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

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

François Le Gland is supervising two others PhD students

- Alexandre Lepoutre, provisional title: *Detection issues in track-before-detect*, université de Rennes 1, started in October 2010, expected defense in 2016, funding: ONERA grant, co-direction: Olivier Rabaste (ONERA, Palaiseau),
- Kersane Zoubert-Ousseni, provisional title: *Particle filters for hybrid indoor navigation with smart-phones*, université de Rennes 1, started in December 2014, expected defense in 2017, funding: CEA grant, co-direction: Christophe Villien (CEA LETI, Grenoble).

Valérie Monbet is supervising one other PhD student

- 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 October 2018, co-direction : Jean-François Dupuy (INSA de Rennes), Laurent Rouvière (université de Haute Bretagne).

### 8.2.3. Thesis committees

François Le Gland has been a reviewer for the PhD theses of Jana Kalawoun (université Paris Sud, Orsay, advisors: Gilles Celeux and Patrick Pamphile) and Antoine Campi (université Paul Sabatier, Toulouse, advisors: Christophe Baehr, Alain Dabas and Pierre Del Moral). He has also been a member of the committee for the PhD thesis of Eugenia Koblentz (Universidad Carlos III de Madrid, advisor: Joaquín Míguez).

Valérie Monbet has been a member of the committee for the PhD theses of Xavier Kergadallan (École des Ponts ParisTech, advisor: Michel Benoit) and Khalil El Waled (université de Haute Bretagne, advisor: Dominique Dehay).

## 8.3. 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 in the bibliography, members of ASPI have also given the following presentations.

Frédéric Cérou has presented the results about the convergence of ABC at the probability and stochastic processes seminar of université de Rennes 1, and at the applied mathematics seminar of université de Nantes, both in November 2015.

Patrick Héas has given a talk on 3D wind field reconstruction by infrared sounding, at EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) in Darmstadt, Germany, in June 2015, and a talk on reduced-order modeling of hidden dynamics, at the international workshop on reduced basis, POD and PGD model reduction techniques, held in Cachan in November 2015.

François Le Gland has given a talk on simulation-based algorithms for the optimization of sensor deployment at the department of signal theory and communications of Universidad Carlos III de Madrid, in February 2015, and a talk on marginalization in rare event simulation for switching diffusions at the ONERA workshop on particle algorithms, held in Toulouse in May 2015.

Valérie Monbet has given a talk on switching autoregressive models for stochastic weather generators, and application to temperature series, at the kick-off meeting of the SEACS project, held in Landeda in May 2015.

Kersane Zoubert-Oussenin has given a poster presentation at the summer school on **Foundations and Advances in Stochastic Filtering**, held in Barcelona in June 2015.

## 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° 4, pp. 2034–2040, <http://dx.doi.org/10.1109/TIT.2010.2040857>
- [3] 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>
- [4] F. CÉROU, A. GUYADER. *Adaptive multilevel splitting for rare event analysis*, in "Stochastic Analysis and Applications", March 2007, vol. 25, n° 2, pp. 417–443, <http://dx.doi.org/10.1080/07362990601139628>

- [5] 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
- [6] T. FURON, A. GUYADER, F. CÉROU. *On the design and optimization of Tardos probabilistic fingerprinting codes*, in "10th International Workshop on Information Hiding, Santa Barbara", Berlin, K. SOLANKI, K. SULLIVAN, U. MADHOW (editors), Lecture Notes in Computer Science, Springer, May 2008, vol. 5284, pp. 341–356, [http://dx.doi.org/10.1007/978-3-540-88961-8\\_24](http://dx.doi.org/10.1007/978-3-540-88961-8_24)
- [7] 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
- [8] F. LE GLAND, L. MEVEL. *Exponential forgetting and geometric ergodicity in hidden Markov models*, in "Mathematics of Control, Signals, and Systems", 2000, vol. 13, n<sup>o</sup> 1, pp. 63–93, <http://dx.doi.org/10.1007/PL00009861>
- [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<sup>o</sup> 337, pp. 351–389, [http://dx.doi.org/10.1007/11587392\\_11](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](http://dx.doi.org/10.1007/978-1-4757-3437-9_12)

## Publications of the year

### Articles in International Peer-Reviewed Journals

- [11] P. AILLIOT, D. ALLARD, V. MONBET, P. NAVEAU. *Stochastic weather generators: an overview of weather type models*, in "Journal de la Société Française de Statistique", 2015, vol. 156, n<sup>o</sup> 1, pp. 101–113, <https://hal.archives-ouvertes.fr/hal-01167055>
- [12] P. AILLIOT, J. BESSAC, V. MONBET, F. PÈNE. *Non-homogeneous hidden Markov-switching models for wind time series*, in "Journal of Statistical Planning and Inference", 2015, vol. 160, pp. 75–88 [DOI : 10.1016/J.JSPI.2014.12.005], <https://hal.archives-ouvertes.fr/hal-00974716>
- [13] J.-D. ALBERT, V. MONBET, A. JOLIVET-GOUGEON, N. FATIH, M. LE CORVEC, M. SECK, F. CHARPENTIER, G. COIFFIER, C. BOUSSARD-PLÉDEL, B. BUREAU, P. GUGGENBUHL, O. LORÉAL. *A novel method for a fast diagnosis of septic arthritis using mid infrared and deported spectroscopy*, in "Joint Bone Spine", 2016, forthcoming [DOI : 10.1016/J.JBSPIN.2015.05.009], <https://hal-univ-rennes1.archives-ouvertes.fr/hal-01243032>
- [14] J. BESSAC, P. AILLIOT, V. MONBET. *Gaussian linear state-space model for wind fields in the North-East Atlantic*, in "Environmetrics", February 2015, vol. 26, n<sup>o</sup> 1, pp. 29–38 [DOI : 10.1002/ENV.2299], <https://hal.inria.fr/hal-01100142>

- [15] G. BIAU, F. CÉROU, A. GUYADER. *New insights into Approximate Bayesian Computation*, in "Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques", February 2015, vol. 51, n<sup>o</sup> 1, pp. 376-403 [DOI : 10.1214/13-AIHP590], <https://hal.archives-ouvertes.fr/hal-00721164>
- [16] A. GUYADER, N. HENGARTNER, N. JÉGOU, E. MATZNER-LØBER. *Iterative Isotonic Regression*, in "ESAIM: Probability and Statistics", 2015, vol. 19, pp. 1-23 [DOI : 10.1051/PS/2014012], <https://hal.archives-ouvertes.fr/hal-00832863>
- [17] D. JACQUEMART, J. MORIO, F. LE GLAND. *Some ideas for bias and variance reduction in the splitting algorithm for diffusion processes*, in "Journal of Computational Science", November 2015, vol. 11, pp. 58-68 [DOI : 10.1016/J.JOCS.2015.09.005], <https://hal.inria.fr/hal-01253763>
- [18] K. LÉON, K. PICHAVANT-RAFINI, H. OLLIVIER, V. MONBET, E. L'HER. *Does Induction Time of Mild Hypothermia Influence the Survival Duration of Septic Rats?*, in "Therapeutic Hypothermia and Temperature Management", June 2015, vol. 5, n<sup>o</sup> 2, pp. 85-88 [DOI : 10.1089/THER.2014.0024], <https://hal.archives-ouvertes.fr/hal-01259883>
- [19] V. MONBET. *Editorial to the special issue on stochastic weather generators*, in "Journal de la Société Française de Statistique", 2015, vol. 156, n<sup>o</sup> 1, pp. 97-98, Numéro spécial : Génération aléatoire de conditions météorologiques, <https://hal.archives-ouvertes.fr/hal-01240896>
- [20] R. PASTEL, J. MORIO, F. LE GLAND. *Extreme density level set estimation for input-output functions via the adaptive splitting technique*, in "Journal of Computational Science", January 2015, vol. 6, pp. 40-46 [DOI : 10.1016/J.JOCS.2014.11.001], <https://hal.inria.fr/hal-01110391>

### International Conferences with Proceedings

- [21] P. BUI QUANG, C. MUSSO, F. LE GLAND. *The Kalman Laplace filter: a new deterministic algorithm for nonlinear Bayesian filtering*, in "Proceedings of the 18th International Conference on Information Fusion, Washington DC 2015", Washington DC, United States, July 2015, pp. 1657-1663, <https://hal.inria.fr/hal-01183413>

### Conferences without Proceedings

- [22] P. HÉAS, C. HERZET. *Inverse Reduced-Order Modeling*, in "Reduced Basis, POD and PGD Model Reduction Techniques", Cachan, France, November 2015, <https://hal.inria.fr/hal-01245051>

### Scientific Books (or Scientific Book chapters)

- [23] Y. KENNÉ, F. LE GLAND, C. MUSSO, S. PARIS, Y. GLEMAREC, E. VASTA. *Simulation-based algorithms for the optimization of sensor deployment*, in "Proceedings of the 3rd International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences, Metz 2015", H. A. L. THI, T. P. DINH, N. T. NGUYEN (editors), Advances in Intelligent Systems and Computing, Springer, May 2015, vol. 360, pp. 261-272 [DOI : 10.1007/978-3-319-18167-7\_23], <https://hal.inria.fr/hal-01183819>
- [24] P. TANDEO, P. AILLIOT, J. RUIZ, A. HANNART, B. CHAPRON, A. CUZOL, V. MONBET, R. EASTON, R. FABLET. *Combining analog method and ensemble data assimilation: application to the Lorenz-63 chaotic system*, in "Machine Learning and Data Mining Approaches to Climate Science: proceedings of the 4th International Workshop on Climate Informatics", Springer, 2015, pp. 3 - 12 [DOI : 10.1007/978-3-319-17220-0\_1], <https://hal.archives-ouvertes.fr/hal-01202496>



## Other Publications

- [25] P. AILLIOT, B. DELYON, V. MONBET, M. PREVOSTO. *Dependent time changed processes with applications to nonlinear ocean waves*, 2015, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01219717>
- [26] J. BESSAC, P. AILLIOT, J. CATTIAUX, V. MONBET. *Comparison of hidden and observed regime-switching autoregressive models for (u,v)-components of wind fields in the Northeast Atlantic*, January 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01250353>
- [27] P. HÉAS, A. DRÉMEAU, C. HERZET. *An Efficient Algorithm for Video Super-Resolution Based On a Sequential Model*, December 2015, 37 pages, <https://hal.inria.fr/hal-01158551>
- [28] P. HÉAS, C. HERZET. *Reduced-Order Modeling Of Hidden Dynamics*, December 2015, working paper or preprint, <https://hal.inria.fr/hal-01246074>
- [29] V. MONBET, P. AILLIOT. *Sparse vector Markov switching autoregressive models Application to multiple time series of air temperature*, January 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01250058>

## References in notes

- [30] 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>
- [31] 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<sup>o</sup> 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), pp. 174–188, <http://dx.doi.org/10.1109/78.978374>
- [32] 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<sup>o</sup> 5 (Special issue on Large-Scale Dynamic Systems), pp. 899–924, <http://dx.doi.org/10.1109/JPROC.2007.893250>
- [33] O. CAPPÉ, É. MOULINES, T. RYDÉN. *Inference in hidden Markov models*, Springer Series in Statistics, Springer–Verlag, New York, 2005, <http://dx.doi.org/10.1007/0-387-28982-8>
- [34] 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<sup>o</sup> 1, pp. 19–67
- [35] 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>
- [36] 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>

- [37] P. DEL MORAL, A. GUIONNET. *On the stability of interacting processes with applications to filtering and genetic algorithms*, in "Annales de l'Institut Henri Poincaré, Probabilités et Statistiques", 2001, vol. 37, n<sup>o</sup> 2, pp. 155–194
- [38] 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>
- [39] R. DOUC, C. MATIAS. *Asymptotics of the maximum likelihood estimator for general hidden Markov models*, in "Bernoulli", June 2001, vol. 7, n<sup>o</sup> 3, pp. 381–420
- [40] 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<sup>o</sup> 5, pp. 2344–2376, <http://dx.doi.org/10.1214/07-AOS514>
- [41] R. DOUC, É. MOULINES, D. S. STOFFER. *Nonlinear time series: Theory, methods and applications with R examples*, Texts in Statistical Science, Chapman & Hall / CRC Press, Boca Raton, 2014, <http://www.crcpress.com/product/isbn/9781466502253>
- [42] E. B. DYNKIN, R. J. VANDERBEI. *Stochastic waves*, in "Transactions of the American Mathematical Society", February 1983, vol. 275, n<sup>o</sup> 2, pp. 771–779, <http://dx.doi.org/10.2307/1999052>
- [43] 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<sup>o</sup> 3, pp. 24–33
- [44] 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
- [45] D. FRENKEL, B. SMIT. *Understanding molecular simulation. From algorithms to applications*, Computational Science Series, 2nd, Academic Press, San Diego, 2002, vol. 1
- [46] 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>
- [47] P. GLASSERMAN, P. HEIDELBERGER, P. SHAHABUDDIN, T. ZAJIC. *Multilevel splitting for estimating rare event probabilities*, in "Operations Research", July–August 1999, vol. 47, n<sup>o</sup> 4, pp. 585–600, <http://dx.doi.org/10.1287/opre.47.4.585>
- [48] E. GOBET, S. MENOZZI. *Stopped diffusion processes: Boundary corrections and overshoot*, in "Stochastic Processes and their Applications", February 2010, vol. 120, n<sup>o</sup> 2, pp. 130–162, <http://dx.doi.org/10.1016/j.spa.2009.09.014>
- [49] 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<sup>o</sup> 2, pp. 107–113, <http://dx.doi.org/10.1049/ip-f-2.1993.0015>



- [50] F. GUSTAFSSON, F. GUNNARSSON, N. BERGMAN, U. FORSELL, 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<sup>o</sup> 2 (Special issue on Monte Carlo Methods for Statistical Signal Processing), pp. 425–437, <http://dx.doi.org/10.1109/78.978396>
- [51] M. ISARD, A. BLAKE. CONDENSATION — *Conditional density propagation for visual tracking*, in "International Journal of Computer Vision", August 1998, vol. 29, n<sup>o</sup> 1, pp. 5–28, <http://dx.doi.org/10.1023/A:1008078328650>
- [52] M. R. JAMES, F. LE GLAND. *Consistent parameter estimation for partially observed diffusions with small noise*, in "Applied Mathematics & Optimization", July/August 1995, vol. 32, n<sup>o</sup> 1, pp. 47–72, <http://dx.doi.org/10.1007/BF01189903>
- [53] M. JOANNIDES, F. LE GLAND. *Small noise asymptotics of the Bayesian estimator in nonidentifiable models*, in "Statistical Inference for Stochastic Processes", 2002, vol. 5, n<sup>o</sup> 1, pp. 95–130, <http://dx.doi.org/10.1023/A:1013737907166>
- [54] 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<sup>o</sup> 1, pp. 1–25, <http://www.jstor.org/stable/1390750>
- [55] Y. A. KUTOYANTS. *Identification of dynamical systems with small noise*, Mathematics and its Applications, Kluwer Academic Publisher, Dordrecht, 1994, vol. 300
- [56] H. R. KÜNSCH. *Recursive Monte Carlo filters : Algorithms and theoretical analysis*, in "The Annals of Statistics", October 2005, vol. 33, n<sup>o</sup> 5, pp. 1983–2021, <http://www.jstor.org/stable/3448632>
- [57] 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<sup>o</sup> 2 (Special issue honoring Perwez Shahabuddin), Article 9, <http://dx.doi.org/10.1145/1225275.1225280>
- [58] L. LE CAM. *Asymptotic methods in statistical decision theory*, Springer Series in Statistics, Springer-Verlag, New York, 1986
- [59] 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<sup>o</sup> 2, pp. 279–316, [http://dx.doi.org/10.1016/S0304-4149\(03\)00041-3](http://dx.doi.org/10.1016/S0304-4149(03)00041-3)
- [60] 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<sup>o</sup> 1, pp. 144–187, <http://dx.doi.org/10.1214/aoap/1075828050>
- [61] F. LE GLAND, B. WANG. *Asymptotic normality in partially observed diffusions with small noise : application to FDI*, in "Workshop on Stochastic Theory and Control, University of Kansas 2001. In honor of Tyrone E. Duncan on the occasion of his 60th birthday", Berlin, B. PASIK-DUNCAN (editor), Lecture Notes in Control and Information Sciences, Springer-Verlag, 2002, n<sup>o</sup> 280, pp. 267–282
- [62] J. S. LIU. *Monte Carlo strategies in scientific computing*, Springer Series in Statistics, Springer-Verlag, New York, 2001

- [63] B. RISTIĆ, M. S. ARULAMPALAM, N. J. GORDON. *Beyond the Kalman Filter : Particle Filters for Tracking Applications*, Artech House, Norwood, MA, 2004
- [64] 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
- [65] C. J. STONE. *Consistent nonparametric regression (with discussion)*, in "The Annals of Statistics", July 1977, vol. 5, n<sup>o</sup> 4, pp. 595–645, <http://www.jstor.org/stable/2958783>
- [66] S. THRUN, W. BURGARD, D. FOX. *Probabilistic robotics*, Intelligent Robotics and Autonomous Agents, The MIT Press, Cambridge, MA, 2005
- [67] D. VILLEMONAIS. *General approximation method for the distribution of Markov processes conditioned not to be killed*, in "ESAIM: Probability and Statistics", 2014, vol. 18, pp. 441–467, <http://dx.doi.org/10.1051/ps/2013045>
- [68] A. W. VAN DER VAART, J. A. WELLNER. *Weak convergence and empirical processes*, Springer Series in Statistics, Springer-Verlag, Berlin, 1996