INRIA, Evaluation of Theme
Stochastic Methods and Models

Project-team ASPI

March 2010

Project-team title:
Applications of Interacting Particle Systems to Statistics

Scientific leader: François Le Gland

Research center: INRIA Rennes — Bretagne Atlantique

1 Personnel

Personnel (January 2005, creation)

	Misc.	INRIA	CNRS	University	**Total**
DR / Professors		1			**1**
CR / Assistant Professors		2		1	**3**
Permanent Engineers					**0**
Temporary Engineers					**0**
PhD Students				1	**1**
Post-Doc.					**0**
Total	**0**	**3**	**0**	**2**	**5**
External Collaborators					**0**
Visitors (> 1 month)					**0**

Personnel (March 2010, evaluation)

	Misc.	INRIA	CNRS	University	**Total**
DR / Professors		1			**1**
CR / Assistant Professors		1		1	**2**
Permanent Engineers					**0**
Temporary Engineers					**0**
PhD Students	4	1	1		**6**
Post-Doc.		1			**1**
Total	**4**	**4**	**1**	**1**	**10**
External Collaborators					**0**
Visitors (> 1 month)					**0**

Changes in staff

DR / Professors CR / Assistant Professors	Misc.	INRIA	CNRS	University	**Total**
Arrival			1		**1**
Leaving		1	1		**2**

Comments: Jean–Pierre Le Cadre, DR CNRS, joined ASPI in January 2009 and deceased in July 2009.

Current composition of the project-team (March 2010):

- Frédéric Cérou, CR INRIA
- Arnaud Guyader, MC, université Rennes 2
- François Le Gland, DR INRIA
- Anindya Goswami, post-doc
- Adrien Ickowicz, PhD student
- Nordine El Baraka, PhD student, Thalès Communications, Massy
- Rudy Pastel, PhD student, ONERA/DPRS Châtillon
- Paul Bui–Quang, PhD student, ONERA/DTIM Châtillon
- Renaud Cariou, PhD student, DGA/CELAR Bruz
- Sébastien Beyou, PhD student (jointly with EPI FLUMINANCE)

Current position of former project-team members (including PhD students during the 2005–2010 period):

- Fabien Campillo, CR INRIA, moved to Montpellier (EPI MERE) in July 2007
- Jean–Pierre Le Cadre, DR CNRS, deceased in July 2009
- Natacha Caylus, PhD student, deceased in January 2006
- Vu–Duc Tran, PhD student, now associate professor, department of science and technology, Hoa Sen University, Ho Chi–Minh City
- Francis Céleste, PhD student, engineer DGA/CEP, Arcueil
- Julien Guillet, ingénieur–expert, now engineer IPSIS, Cesson–Sévigné
- Liyun He, ingénieur–expert, now PhD student, Télécom Bretagne, Brest
- Vivien Rossi, post–doc, now CR CIRAD, Montpellier
- Cyrille Jégourel, ingénieur–expert
- Rui Han, ingénieur–expert
- Fida El Haje Hussein, post–doc

Last INRIA enlistments:

No INRIA enlistment ever.

Other comments:

Pierre Del Moral has been an external collaborator of ASPI from June 2005 to September 2007, when he moved from université de Nice to INRIA Bordeaux — Sud Ouest.

2 Work progress

2.1 Keywords

sequential Monte Carlo (SMC) method, interacting particle system, Bayesian filtering, particle filtering, localisation, navigation and tracking, sequential data assimilation, rare event simulation, multilevel splitting, fingerprinting and watermarking, global optimization, machine learning, functional data analysis, functional nearest neighbor classification, regression estimation, non–parametric statistics, recommendation system.

2.2 Context and overall goal of the project

The scientific objectives of the ASPI team, as exposed in the proposal submitted in March 2004, are the design, practical implementation and mathematical analysis of interacting Monte Carlo methods, also known as particle methods or sequential Monte Carlo (SMC) methods, with focus on

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

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* the environment by evaluating a fitness function,
- and *interact* so that only the most successful particles (in view of the value 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. The whole context is multidisciplinary, not only because of the many scientific and engineering areas in which these 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.

The ASPI team essentially carries methodological research activities, with the objective to obtain generic results with high potential for applications, and to implement these results on a few practical examples, through continuing collaboration and partnership with public research institutions or with industrial partners, such as

DGA (defence agency), ONERA and EDF RD.

Occasional collaboration during the evaluation period has involved Thalès Alénia Space, Thalès Communications and France Télécom RD (now Orange Labs). Applications currently or recently considered are

> geolocalization and tracking of mobile terminals, terrain–aided navigation, information fusion for indoor localisation, detection in sensor networks, risk assessment in air traffic management, protection of digital documents, option pricing, and credit risk estimation.

2.3 Objectives for the evaluation period

The following objectives have been listed in the proposal submitted in March 2004:

1. Methodology of particle methods.
2. Simulation of rare events.
3. Particle filtering and statistics of hidden Markov models.

As already mentioned in the previous section, 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* the environment by evaluating a fitness function,
- and *interact* so that only the most successful particles (in view of the value 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, ubiquitous computing and ambient intelligence, sensor networks, etc., each particle represents a possible hidden state, and is multiplied 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 special case of splitting or branching methods for the estimation of the small probability of a rare but critical event, for instance the probability that a random process reaches a critical region of the state space, which is a crucial issue in industrial areas such as nuclear power plants, telecommunication networks, protection of digital documents, finance and insurance industry, defence systems, air traffic management, etc. intermediate less critical regions are introduced, and trajectories hitting an intermediate region, i.e. going potentially towards the most critical region, are given offsprings, thus increasing the number of trajectories that eventually hit the most critical region. The same strategy also applies in static cases, for the estimation of the small probability that a random variable, defined as the output of some (large, complex, costly to execute) numerical code with random input, exceeds some extreme level.

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, which provides a very powerful mathematical framework for the mathematical analysis of practical algorithms, mostly in the large sample asymptotics, i.e. as the number of particles grows to infinity. Issues to be considered here are:

> how to resample, when to resample, which Markov model and selection functions to use (as in importance sampling), how many particles to use, etc.,

and what is the impact of these different designs on the approximation.

An additional topic, that was not present in the initial list of objectives, has emerged during the period:

4. Functional data analysis and classification.

2.4 Objective 1: Methodology of particle methods (executive summary)

In mathematical terms, many of the problems mentioned above can be formulated as the problem of evaluating mathematical expectations of the form

$$\langle \gamma_n, \phi \rangle = \mathbb{E}[\phi(X_n) \prod_{k=0}^{n} g_k(X_k)] \qquad \text{and} \qquad \langle \mu_n, \phi \rangle = \frac{\langle \gamma_n, \phi \rangle}{\langle \gamma_n, 1 \rangle} \ ,$$

for *any* function ϕ, which implicitly define γ_n and μ_n as an unnormalized and a normalized (probability) distribution, respectively. Here X_k is a Markov chain taking values in E and characterized by the initial probability distribution η_0 and by the transition kernels Q_k, and g_k are nonnegative potential functions. A naive implementation of Monte Carlo methods would consist in sampling N independent trajectories of the Markov chain, evaluate the function ϕ at the terminal point of the trajectory, and weight the sample by the product of the potential functions evaluated along the trajectory. This results in serious *degeneracy* problems, since the product of positive numbers usually gets exponentially small or large. The major breakthrough has been to exploit the sequential nature of the problem and at each time step to consider using the weights to select the most interesting trajectories and discard the other trajectories, thus introducing some kind of interaction between trajectories. Indeed, the following recurrent equation

$$\mu_{k-1} \xrightarrow{\text{mutation}} \eta_k = \mu_{k-1}\, Q_k \xrightarrow{\text{weighting}} \mu_k = g_k \cdot \eta_k$$

is easily obtained, and the idea is to look for an approximation of the form

$$\mu_k \approx \mu_k^N = \sum_{i=1}^{N} w_k^i \, \delta_{\xi_k^i} \qquad \text{with} \qquad \sum_{i=1}^{N} w_k^i = 1 \ ,$$

in terms of the weighted empirical probability distribution associated with a system of particles, characterized by the particle positions and the particle (nonnegative) weights. Injecting the approximation μ_{k-1}^N into the mutation step yields a finite mixture probability distribution which is not of the desired form, but can be approximated as

$$\mu_{k-1}^N \, Q_k \approx \eta_k^N = \sum_{i=1}^{N} v_k^i \, \delta_{\xi_k^i} \qquad \text{with} \qquad \sum_{i=1}^{N} v_k^i = 1 \ .$$

Possible choices for the new particle positions and the new particle weights include conservation of the weights or multinomial resampling, and it is also possible to switch automatically between these two extreme schemes, or to design others. The first scheme is just a sequential rephrasing of the naive non–interacting Monte Carlo method (also known as sequential importance sampling), in which the weights are just accumulated along trajectories, whereas in the second scheme the weights are used in a different way, and contribute to the selection mechanism. Injecting the approximation η_k^N in the weighting step yields

$$g_k \cdot \eta_k^N = \mu_k^N = \sum_{i=1}^N w_k^i \, \delta_{\xi_k^i} \qquad \text{with} \qquad w_k^i = v_k^i \, g_k(\xi_k^i) \; / \; [\sum_{j=1}^N v_k^i \, g_k(\xi_k^i)] \; ,$$

which is of the desired form, without any further approximation.

Many asymptotic results have been obtained or are currently under investigation, to assess the performance of these numerical approximation schemes as the sample size N increases, i.e. to precise how do $\langle \mu_n^N, \phi \rangle$ and $\langle \gamma_n^N, 1 \rangle$ converge to $\langle \mu_n, \phi \rangle$ and $\langle \gamma_n, 1 \rangle$, respectively or in which sense should the symbol $\approx$ be understood:

> convergence in p–th mean, uniformly over time or over a class of functions, central limit theorem, propagation of chaos, probability of deviation, etc.

Many of these results have been obtained by Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest), an external collaborator of ASPI from June 2005 until September 2007, and can be found in [O6] and in his more recent papers. Yet, ASPI has also contributed and is still contributing on its own to this theory.

2.4.1 Personnel

permanent researchers: Fabien Campillo (until July 2007), Frédéric Cérou, Arnaud Guyader, François Le Gland
external collaborator: Pierre Del Moral (from June 2005 to September 2007)
post–doc: Vivien Rossi

2.4.2 Project–team positioning

Mathematical (asymptotic) analysis of particle methods seems to be a French speciality, with main contributors Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest), Éric Moulines, Randal Douc and Olivier Cappé (Télécom Paris), Nicolas Chopin (ENSAE), Arnaud Doucet (University of British Columbia), Christophe Andrieu (University of Bristol). Outside the French community, other contributors to the theory include Dan Crişan and Ajay Jasra (Imperial College), Hans–Rüdi Künsch (ETH Zürich).

Within INRIA, and besides EPI ALEA, INRIA Bordeaux — Sud Ouest, Monte Carlo methods are investigated in EPI TOSCA, INRIA Sophia Antipolis — Méditerranée and INRIA Nancy — Grand Est, or in EPI MATHFI, INRIA Paris — Rocquencourt.

2.4.3 Scientific achievements

Nonasymptotic variance This is a collaboration with Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest). A non asymptotic theorem for interacting particle approximations of unnormalized Feynman–Kac distributions has been obtained [R6]. An original stochastic analysis has been developed, based on Feynman–Kac semigroup techniques combined with recently developed coalescent tree–based functional representations of particle block distributions. Regularity conditions have been introduced under which the $\mathbb{L}_2$–relative error of these weighted particle distributions grows linearly with respect to the time horizon, yielding what seems to be the first results of this type for this class of unnormalized distributions. These results have also been illustrated in the context of particle simulation of static Boltzmann–Gibbs distributions and restricted distributions, with a special interest in rare event analysis. For rare events, these results yield asymptotic efficiency when used in conjunction with a good importance function.

Adaptive sample size and non–extinction This is a collaboration with Nadia Oudjane (EDF RD, department OSIRIS) In the special case of a Feynman–Kac distribution, where the selection functions can possibly take the zero value, which may occur in many important practical situations

- simulation of rare events using an *importance splitting* approach, or simulation of a r.v. in the tail of a given probability distribution,
- simulation of a Markov chain conditioned or constrained to visit a given sequence of subspaces of the state space,
- algorithms of approximate nonlinear filtering, somehow related to ABC (approximate Bayesian computation), where hidden state and observation are simulated jointly, and where the simulated observation is validated against the actual observation [O9, O8, O29], e.g. if there is no explicit expression available for the likelihood function, or if there does not even exist a likelihood function (nonadditive observation noise, noise–free observations, etc.),

it can happen that the evaluation of the selection function returns the zero value for all the particles generated at the end of the mutation step, i.e. the particle systems dies out and the algorithm cannot go on. A sequential particle algorithm, already proposed in [O25], [O21], has been studied [B5], [R11], and a central limit theorem has been obtained, with two alternative proofs. By construction, this sequential algorithm automatically keeps the particle system alive, i.e. it ensures its non–extinction, and it can be seen as a *fixed performance* policy, as opposed to the usual nonsequential algorithm which is a *fixed effort* policy.

Combining weighting and resampling A particle approximation has been studied, that combines SIS and SIR algorithms in the sense that only a fraction of the importance weights is used for resampling, and two different approaches have been proposed to analyze its performance [J13]. The starting point is a factorization of the potential functions as the product of an importance weight function and a resampling weight function.

In the first approach, based on a representation in path–space, the importance weight functions appear only in the test function, and it is therefore easy to analyze the joint particle approximation of unnormalized distributions (and normalizing constants and normalized distributions, as a by–product) for the reference model and for several alternate models at the same time, just by choosing the appropriate test function in the central limit theorem. In other words, this first approach is appropriate to analyze particle approximations in statistical models depending on some parameter, in sensitivity analysis, etc.

In the second approach, based on a representation in terms of a multiplicative functional, importance weights that are not used for resampling are treated as particles. This interpretation is of independent interest and could be used to analyze particle approximation with adaptive resampling schemes, see below, where the decision to use resampling weights only vs. importance weights only is made dependent on an empirical criterion (effective number of particles, entropy of the sample, etc.) evaluated using the current particle approximation of the unnormalized distribution. The idea is simply to introduce a factorization of the potential functions that depends on the current unnormalized distribution: this results in a representation in terms of a McKean model, and the associated particle approximation could then be easily analyzed.

Adaptive resampling This is a collaboration with Élise Arnaud (université Joseph Fourier and EPI PERCEPTION, INRIA Grenoble — Rhône Alpes). A longstanding problem in particle or sequential Monte Carlo (SMC) methods is to mathematically prove the popular belief that resampling does improve the performance of the estimation (this of course is not always true, and the real question is to clarify classes of problems where resampling helps). A more pragmatic answer to the problem is to use adaptive procedures that have been proposed on the basis of heuristic considerations, where resampling is performed only when it is felt necessary, i.e. when some criterion (effective number of particles, entropy of the sample, etc.) reaches some prescribed threshold. It still remains to mathematically prove the efficiency of such adaptive procedures. The first contribution has been to consider a design where resampling is performed at some intermediate fixed time instants, and to optimize the asymptotic variance of the estimation error w.r.t. the resampling time instants. The second contribution [C1] has been to prove a central limit theorem for particle methods with adaptive resampling, using an interpretation of particle methods where importance weights are treated as particles [J13], as long as they are not used for resampling purpose, and to minimize the asymptotic variance w.r.t. the threshold.

Non–parametric learning of the optimal importance distribution This is a collaboration with Nadia Oudjane (EDF RD, department OSIRIS). Evaluating an integral or a mathematical expectation of a nonnegative function can always be seen as computing the normalization constant in a Boltzmann–Gibbs probability distribution. When the probability distribution and the nonnegative function do not agree, i.e. have significant contributions in different parts of the integration space, then the variance of the estimator can be very large, and one should use another importance distribution, ideally the optimal (zero variance) importance distribution μ^* which unfortunately cannot be used since it depends on the desired (but unknown) integral. Alternatively, sequential methods have been designed (under different names, such as annealed sampling, progressive correction, multilevel splitting, etc., depending on the context) which not only provide an expression for the desired integral as the product of intermediate normalization constants, but ultimately provide as well an N–sample approximately distributed according to the optimal importance distribution. From the weighted empirical probability distribution associated with this sample, a regularized probability distribution μ_N can be obtained, using a kernel method or a simple histogram, and can be used as an almost optimal importance distribution to estimate the original integral with a M–sample distributed according to μ_N. The variance of the resulting estimator depends on the product of the inverse sample size $1/M$ by the χ^2–distance between the almost optimal importance distribution μ_N and the optimal (zero variance) importance distribution μ^*.

The contribution has been to provide an estimate of this χ^2–distance, under mild assumptions. The impact of dimension on density estimation is a limiting factor here, but the variance reduction is very significant in moderate dimensions. A further work would be to compare this non–parametric approach with a parametric approach, e.g. based on the cross–entropy method.

2.4.4 Collaborations

Collaboration with Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest) on nonasymptotic variance estimation, with Élise Arnaud (université Joseph Fourier and EPI PERCEPTION, INRIA Grenoble — Rhône Alpes) on adaptive resampling and with Nadia Oudjane (EDF RD, department OSIRIS) on adaptive sample size and non–extinction, and on non–parametric learning of the optimal importance distribution.

2.4.5 External support

No external support has been applied for.

2.4.6 Self assessment

What makes ASPI (and other French groups) visible on the international scene, is not only its applications of particle methods, which many other groups throughout the world do very well within their own scientific domain, but also its rigorous mathematical analysis of some of the most popular particle algorithms.

2.5 Objective 2: Simulation of rare events (executive summary)

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

> nuclear power plants, telecommunication networks, protection of digital documents, finance and insurance industry, defence systems, 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 [O22], 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 [J6], the Feynman–Kac formalism 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 thresholds. This problem is characterized by the lack of an analytical expression for the function F, 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.

Rare event simulation Multilevel splitting can also be used in dynamical situations. 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).

2.5.1 Personnel

permanent researchers: Frédéric Cérou, Arnaud Guyader, François Le Gland
external collaborator: Pierre Del Moral (from June 2005 to September 2007)
PhD student: Rudy Pastel
ingénieur–expert: Cyrille Jégourel
post–doc: Fida El Haje Hussein, Anindya Goswami

2.5.2 Project-team positioning

The question of rare event simulation is very large and occurs in many different domains, and ASPI is still a newcomer, hence the picture given here is certainly incomplete. However, important contributors definitely include Paul Glasserman (Columbia University) for his early mathematical analysis of multilevel splitting, Reuven Rubinstein (Technion) and Dirk Kroese (University of Queensland) for their introduction of the cross–entropy method, and Paul Dupuis (Brown University) for his construction of good, if not optimal, importance distributions (for importance sampling) and importance functions (for multilevel splitting), using large deviations arguments. Within INRIA, Bruno Tuffin (EPI DIONYSOS, INRIA Rennes — Bretagne Atlantique) is also a very active contributor, especially through collaboration with Pierre L'Écuyer (université de Montréal).

2.5.3 Scientific achievements

During the evaluation period, a major application area of rare event simulation was the reliability of protection mechanisms for digital documents, i.e. the evaluation of (small) probabilities of false alarm in watermarking or fingerprinting schemes. This activity has started on the occasion of the ANR project NEBBIANO, within the SETIN (Sécurité et Informatique) programme.

Design and optimization of Tardos probabilistic fingerprinting codes This is a collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique). Gábor Tardos [O34] was the first to give a construction of a fingerprinting code whose length meets the lowest known bound. This was a real breakthrough because the construction is very simple. Its efficiency comes from its probabilistic nature. However, although Tardos almost gave no argument of his rationale, many parameters of his code are precisely fine–tuned. We propose [B2] this missing rationale supporting the code construction. The key idea is to make the statistics of the scores as independent as possible from the collusion process. Tardos optimal parameters are rediscovered. This interpretation allows small improvements when some assumptions hold on the collusion process.

Estimating the minimal length of Tardos code This is a collaboration with Luis Pérez–Feire from GRADIANT (Galician R&D Center in Advanced Telecommunications, Vigo, Spain) and with Teddy Furon from INRIA Rennes Bretagne Atlantique (project–team TEMICS). We are interested in the minimal length of a binary probabilistic traitor tracing code. We consider the code construction proposed by Gábor Tardos in 2003, with the symmetric accusation function as improved by Boris Škorić et al. The length estimation is based on two pillars. First, we consider the worst case attack that a group of c colluders can lead. This attack minimizes the mutual information between the code sequence of a colluder and the pirated sequence. Second, an algorithm pertaining to the field of rare event analysis is constructed in order to estimate the probabilities of error: the probability that an innocent user is framed, and the probabilities that all colluders are missed. Therefore, for a given collusion size, we are able to estimate the minimal length of the code satisfying some error probability constraints [B3]. This estimation is far lower than the known lower bounds.

Experimental assessment of reliability for watermarking and fingerprinting This is a collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique). The concept of reliability in watermarking has been introduced [J7] as the ability of assessing that a probability of false alarm is very low and below a given significance level. We propose an iterative and adaptive algorithm which estimates very low probabilities of error. It performs much quicker and more accurately than a classical Monte Carlo estimator. The article finishes with applications to zero-bit watermarking (probability of false alarm, error exponent), and to probabilistic fingerprinting codes (probability of wrongly accusing a given user, code length estimation). Prior to publication, a patent [P2] has been jointly submitted in August 2008 by INRIA and Bretagne Valorisation.

Estimating the probability of false alarm for zero–bit watermarking This again is a collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique). Assessing that a probability of false alarm is below a given significance level is a crucial issue in watermarking. We have proposed an iterative and adaptive algorithm which estimates very low probabilities of error. Some experimental investigations validate its performance for a rare detection scenario where there exists a close form formula of the probability of false alarm. Our algorithm appears to be much quicker and more accurate than a classical Monte Carlo estimator. It even allows the experimental measurement of error exponents [C12].

The remaining part is devoted to presenting scientific achievements in the methodology of splitting methods.

Characterization of optimal importance function and optimal threshold The 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.

In this way, 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.

The contribution [J6] has been to study the impact of the shape of the intermediate regions (selection of the importance function), of the thresholds (levels), of the population size, on the asymptotic variance, obtained through a central limit theorem. For instance, assuming that the optimal importance function is used, the levels should be selected in such a way that the probability to reach the k–th level before time T, given that the $(k-1)$–th level has been reached before time T, does not depend on k.

Adaptive threshold selection Besides importance sampling, a widespread technique to estimate rare event probabilities is multilevel splitting, which requires at least some knowledge of the system, to decide where to place the intermediate level sets. This is not always possible, and an adaptive algorithm has been proposed to cope with this problem, and has been analyzed thoroughly in the one–dimensional case.

Assuming that the problem is to estimate the probability P that a one–dimensional Markov process eaches the level $M \gg 1$ before returning to the origin, and taking advantage of the findings about optimal threshold selection, the adaptive algorithm fixes beforehand the probability q to reach the next level before returning to the origin. At each stage, N excursions are simulated that try to approach the final level M, and the $k = \lfloor q\, N \rfloor$ excursions approaching the level M most closely are selected, which implicitly defines the next level, and are given offsprings so that the population size N remains constant. The algorithm continues until generation L where the current (virtual) level exceeds the final level M, i.e. L denotes the first generation where at least one excursion reaches the level M, and let k_{L+1} denote the exact number of such excursions. Then, the rare event probability is estimated as

$$P \approx \widehat{P}_N = \frac{k_{L+1}}{N} \, (\frac{k}{N})^L \ ,$$

and the number of (virtual) intermediate levels is L.

In this one–dimensional framework, almost sure convergence and asymptotic normality of the estimator has been proved, with the same variance as other algorithms that use given intermediate levels. It has also been showed on numerical examples that this method can be used for multidimensional problems as well, even if there is still no convergence result in this case.

Rare event simulation for a static distribution This is a collaboration with Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest) and with Teddy Furon (EPI TEMICS, INRIA Rennes — Bordeaux Atlantique). 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 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.

The contribution [R5] has been to study the impact of design parameters, such as the intermediate levels or the Metropolis kernel introduced in the mutation step, on the asymptotic variance obtained through a central limit theorem. This work has been recently extended [R9], in collaboration with Nicolas Hengartner (LANL) and Éric Matzner–Løber (université de Rennes 2).

2.5.4 Collaborations

Collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique) on reliability of watermarking and fingerprinting, with Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest) on rare event simulation for a static distribution, with Gerardo Rubino and Bruno Tuffin (EPI DIONYSOS, INRIA Rennes — Bretagne Atlantique) on splitting methods, with Pascal Lezaud (DGAC/DTI), Henk Blom (NLR) and Jaroslav Krystul (Twente University) on rare event simulation methods adapted to stochastic hybrid systems, with Jérôme Morio and Rudy Pastel (ONERA/DPRS) on extreme quantile estimation, with Mathieu Chouchane and Sébastien Paris (université Paul Cézanne and LSIS) on cross–entropy method for global optimization, with Nadia Oudjane (EDF RD, department OSIRIS) on out of the money option pricing and value–at–risk estimation, with Nicolas Hengartner (LANL) and Éric Matzner–Løber (université de Rennes 2) on Monte Carlo estimation of extreme quantiles of posterior probabilities.

2.5.5 External support

Research on splitting methods is / has been supported by INRIA under the ARC RARE on rare event simulation by Monte Carlo, by ANR under the project NEBBIANO on reliability of protection of digital documents, within the SETIN (Sécurité et Informatique) programme, by the FP6 project iFLY on highly automated air traffic management, within the Aeronautics and Space programme, and by ONERA through a scholarship on rare event simulation for the analysis of complex defence systems.

Applications to global optimization is supported by DGA/CTSN under a contract on optimization of sensors location and activation, by DGA/CEP through a thesis on trajectory optimization for a mobile robot localization, and by DGA/CELAR through a thesis on trajectory optimization to minimze the detection of an aircraft.

2.5.6 Self assessment

Of the four objectives, this is the one where more effort was put during the evaluation period, and yet it could be argued that ASPI is still an outsider in this area. This is in opposition with the objective on particle filtering, in which ASPI has a well–established expertise. It is certainly true that ASPI was a newcomer, but it already has had a major contribution, even though not always clearly understood, in bringing in a very efficient and general–purpose methodology to analyze the performance of already existing or new and original splitting methods, under very general assumptions. Would it be for this single reason only, the effort put here should be continued. Obviously, other applications should be considered as well, among which the related issue of global optimization. This point is further developped in Section 5.1.

2.6 Objective 3: Particle filtering (executive summary)

During the evaluation period, two different application domains of particle filtering have been investigated in ASPI:

- localization, navigation and tracking,
- sequential data assimilation.

The continuing interest in the former domain finds its roots in the well–established expertise of ASPI, and it is worth noting that the increasing interest in the latter domain has started on the occasion of the ARC project ADOQA.

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 [O16, O14], 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 [O15, O28], 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 to several applications of particle filtering in positioning, navigation and tracking, such as geolocalization and tracking in a wireless network, terrain–aided navigation and information fusion for indoor localisation.

2.6.1 Personnel

permanent researchers: Fabien Campillo (until July 2007), François Le Gland
PhD students: Nordine El Baraka, Paul Bui–Quang, Vu–Duc Tran
ingénieurs–experts: Liyun He, Julien Guillet
post–doc: Vivien Rossi

2.6.2 Project-team positioning

Particle filtering and its applications to localization, navigation and tracking, has raised huge interest worldwide, across many scientific domains, such as target tracking, video tracking, mobile robotics, etc, with main contributors Fredrik Gustafsson and his group (Linköping University), Simon Maskell and David Salmond (QinetiQ, Malvern and Farnborough), Neil Gordon (DSTO, Adelaide), Yvo Boers and Hans Driessen (Thalès Nederland, Hengelo), Ba–Ngu Vo (University of Melbourne), Christian Musso (ONERA/DTIM, Châtillon), Petar Djurić (SUNY at Stony Brook), Dieter Fox (University of Washington), Sebastian Thrun (Stanford University). Within INRIA, François Caron (EPI ALEA, INRIA Bordeaux — Sud Ouest) and Patrick Pérez (EPI VISTA, INRIA Rennes — Bretagne Atlantique, now with Technicolor) are / have been very active contributors.

Main contributors in sequential data assimilation and especially ensemble Kalman filtering are Geir Evensen (Hydro Research Centre, Bergen), Laurent Bertino (Nansen Environmental and Remote Sensing Center, Bergen), Peter Jan van Leeeuwen (Utrecht University), Jeffrey L. Anderson and Chris Snyder (National Center for Atmospheric Research, Boulder) and Thomas Bengtsson (Bell Labs, Statistics and Data Mining Department, Murray Hill). Within INRIA, Étienne Mémin (EPI FLUMINANCE, INRIA Rennes — Bretagne Atlantique) is also a very active contributor.

2.6.3 Scientific achievements

During the evaluation period, several applications in localization, navigation and tracking have been studied:

- information fusion for indoor localization (FIL),
- geolocalization and tracking in a wireless network,
- terrain–aided navigation,

which are described in Sections 4.2 and 4.5. The remaining part is devoted to presenting scientific achievements in sequential data assimilation.

Asymptotics of the ensemble Kalman filter This is a collaboration with Valérie Monbet (université de Bretagne Sud). Very little was known about the asymptotic behaviour of the ensemble Kalman filter [O11], whereas on the other hand, the asymptotic behaviour of many different classes of particle filters is well understood, as the number of particles goes to infinity. Interpreting the ensemble elements as a population of particles with mean–field interactions, and not only as an instrumental device producing an estimation of the hidden state as the ensemble mean value, it has been possible to prove the convergence of the ensemble Kalman filter, with a rate of order $1/\sqrt{N}$, as the number N of ensemble elements increases to infinity [B4] and [R10]. In addition, the limit of the empirical distribution of the ensemble elements has been exhibited, which differs from the usual Bayesian filter. Several cases have been investigated, from the simple case where the drift coefficient is bounded and globally Lipschitz continuous, to the more realistic case where the drift coefficient is locally Lipschitz continuous, with polynomial growth. In all these cases, the observation coefficient was assumed linear, so that the analysis step for each ensemble element has exactly the same structure as the analysis step of the usual Kalman filter.

The next step is to study the asymptotic normality of the estimation error, i.e. to prove a central limit theorem. It is somehow expected that the asymptotic variance for the ensemble Kalman filter would be smaller than the known asymptotic variance for the different brands of particle filters, just because the ensemble Kalman filter follows essentially a parametric approach, where only the first two empirical moments are propagated, whereas the particle filters follow a fully nonparametric approach.

Particle filter for mean–field models This is a collaboration with Christophe Baehr (Météo–France, centre national de recherche météorologique). The motivating application is the estimation of Lagrangian velocity in a turbulent flow: to filter out observation noise a Bayesian approach is used, with a simplified Langevin model [O26] as the prior for the Lagrangian velocity. This model involves local means of the Eulerian velocity field, which can be expressed in terms of the probability distribution of the Lagrangian velocity. Other nonlinear terms in the model, such as the mean pressure gradient, the turbulent kinetic energy k and its dissipation rate ε are either considered as unknown random variables, with a somehow arbitrary prior probability distribution, or are related with local Eulerian means and can then be expressed in terms of the probability distribution of the Lagrangian velocity. In other words, the proposed simplified Langevin model is a special example of a nonlinear McKean model with mean–field interactions, where the drift coefficient depends on the probability distribution of the solution.

The original estimation problem reduces to the estimation of the hidden state in a nonlinear Markov model, and numerical approximations have been studied, with two populations of particles: the first population of particles with mean–field interactions learns the unconditional probability distribution of the hidden state, whereas the second population of particles approximates the Bayesian filter, i.e. the conditional probability distribution of the hidden state given the observations.

Alternatively, since noisy observations are available, the local Eulerian means can be expressed in terms of the conditional probability distribution of the Lagrangian velocity given the observations. This results in a much simpler model, where a single population of particles is sufficient to approximate the Bayesian filter.

2.6.4 Collaborations

Collaboration on applications in localization, navigation and tracking with Christian Musso (ONERA/DTIM) and Olivier Rabaste (ONERA/DEMR) on high–dimensional particle filtering and on Monte Carlo methods for the detection of dim targets, respectively.

Collaboration on applications in sequential data assimilation with Étienne Mémin (EPI FLUMINANCE, INRIA Rennes — Bretagne Atlantique), with Valérie Monbet and Anne Cuzol (université de Bretagne Sud), and with Christophe Baehr (Météo–France, centre national de recherche météorologique).

2.6.5 External support

Applications in localization, navigation and tracking have been supported by France Télécom RD (now Orange Labs) under a CRE on localization, by DGA/SPNuM under a PEA project and a CIFRE grant on terrain–aided navigation, by ONERA through a scholarship on high–dimensional particle filtering, by ANR under the project FIL on information fusion for indoor localization, within the Télécommunications programme.

Applications in sequential data assimilation have been supported by INRIA under the ARC ADOQA on data assimilation for air quality, and by ANR under the project PREVASSEMBLE on ensemble methods for prediction and data assimilation, within the COSINUS (conception and simulation) programme.

2.6.6 Self assessment

There is a mature activity, with many applications, in localization, navigation and tracking, which belongs to the well–established expertise of ASPI and which should be continued. Establishing connections with major contributors to particle filtering in mobile robotics (Dieter Fox, Sebastian Thrun) should be very fruitful.

By comparison, ASPI is still an outsider in the area of sequential data assimilation. However, interesting collaborations with Météo–France and with other major contributors have already been established, especially within the ongoing ANR project PREVASSEMBLE, where many interesting problems have been listed that clearly fall within the expertise of ASPI.

2.7 Emerging objective: Functional data analysis (executive summary)

In pattern recognition and statistical learning, also known as machine learning, nearest neighbor (NN) algorithms are amongst the simplest but also very powerful available algorithms. Basically, given a training set of data, i.e. an N–sample of i.i.d. object–feature pairs (X_i, Y_i) for $i = 1, \ldots, N$, with real–valued features, the question is how to generalize, that is to guess the feature Y associated with any new object X, with the same probability distribution as the X_i's. 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 X, for some given metric.

In general, there is no way to guess exactly the value of Y, and the minimal error that can be done is that of the Bayes estimator $\mathbb{E}[Y \mid X]$, which cannot be computed in general by lack of knowledge of the distribution of the pair, but the Bayes estimator can help to characterize the strength of the method. The best that can be obtained is that the NN estimator converges, say when the sample size grows, to the Bayes estimator. This is what has been proved in great generality by Stone [O33] for the mean square convergence, provided that X is a d-dimensional vector, Y is square–integrable, and the ratio k/N goes to 0. The NN estimator is not the only local averaging estimator having this property, but it is arguably the simplest.

In settings for which the estimator is convergent, there is still the question of the rate of convergence, and how to choose the parameter k in order to achieve the best rate of convergence. As noticed by Kulkarni and Posner [O20], the rate of convergence of the NN estimator is closely related to the notion of metric entropy, introduced in the late fifties by Kolmogorov and Tihomirov [O17]. These tools are to be used to study cases and algorithm refinements that are not yet to be found in the literature.

The situation is radically different in general infinite dimensional spaces, when the objects to be classified are functions, images, etc. Issues to be considered here are:

> under which conditions does the NN estimator converge, which rate of convergence can be achieved, how to choose the parameter k (as a function of N), etc.

2.7.1 Personnel

permanent researchers: Frédéric Cérou, Arnaud Guyader

2.7.2 Project-team positioning

Main contributors include Gérard Biau (université Pierre et Marie Curie, ENS Paris and EPI CLASSIC, INRIA Paris — Rocquencourt), Luc Devroye (McGill University), Gábor Lugosi (Pompeu Fabra University), László Györfi (Budapest University of Technology and Economics), Adam Krzyžak (Concordia University).

2.7.3 Scientific achievements

Nearest neighbor classification in infinite dimension The k-nearest neighbor classifier g_N consists in the simple following rule: look at the k nearest neighbors of X in the N–sample, and choose 0 or 1 for its label according to the majority rule. In the case where $(\mathcal{F}, d) = (\mathbb{R}^m, \| \cdot \|)$, this classifier is universally consistent [O33] its probability of error converges to the Bayes risk as N goes to infinity, whatever the joint probability distribution of (X, Y), provided that k/N goes to 0. 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 $\mathcal{F}$ needs to be complete for the metric d, 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 [J8] If the metric space $(\mathcal{F}, d)$ is separable and if some *Besicovich condition* holds then the nearest neighbor classifier is weakly consistent, i.e.

$$\mathbb{P}[g_N(X) \neq Y] \longrightarrow L^* \ ,$$

as $N \uparrow 0$, where L^* denotes the Bayes probability of error, or Bayes risk. Note that the Besicovich condition is always fulfilled in finite dimensional vector spaces (this result is called the Besicovich theorem), and that a counterexample can be given [O27] 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 This is a collaboration with Gérard Biau (université Pierre et Marie Curie, ENS Paris and EPI CLASSIC, INRIA Paris — Rocquencourt). 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 pair (X, Y), have been investigated [J2] when X is in a suitable ball in some functional space. Using compact embedding theory, we present explicit and general finite sample bounds on the expected squared difference between k-nearest neighbor estimator and 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.

Rate of convergence of the bagged nearest neighbor estimator This is a collaboration with Gérard Biau (université Pierre et Marie Curie, ENS Paris and EPI CLASSIC, INRIA Paris — Rocquencourt). Bagging is a simple way to combine estimators in order to improve their performance. This method, suggested by Breiman in 1996, proceeds by resampling from the original data set, constructing a predictor from each subsample, and decide by combining. By bagging an N-sample, the crude nearest neighbor regression estimator is turned into a consistent weighted nearest neighbor regression estimator, which is amenable to statistical analysis. Letting the resampling size k_N grow with N in such a way that $k_N \to \infty$ and $k_N/N \to 0$, it has been shown [J1] that this estimator achieves optimal rate of convergence, independently from the fact that resampling is done with or without replacement. Since the estimator with the optimal rate of convergence depends on the unknown distribution of the observations, adaptation results by data–splitting are also obtained.

2.7.4 Collaborations

Collaboration with Gérard Biau (université Pierre et Marie Curie, ENS Paris and EPI CLASSIC, INRIA Paris — Rocquencourt) on functional nearest neighbor estimation.

2.7.5 External support

No external support was obtained so far.

2.7.6 Self assessment

It is expected that this emerging activity should rapidly find some application domains, e.g. in the statistical analysis of recommendation systems, that would be a source of interesting problems and provide also some external support. A proposition was submitted to the ARC programme: even though it was not selected, it could serve as a basis for a submission to another programme. This point is further developped in Section 5.2.

3 Knowledge dissemination

3.1 Publications

	2006	2007	2008	2009	2010
PhD Thesis				1	2
Journal	2	3	1	2	4
Book chapter	1	1	1	2	1
Conference proceedings	5	4	7	6	1
Technical report	4		2	4	1
Patent			2		

Major journals in the field and number of papers coauthored by members of the project–team that have been accepted during the evaluation period.

1. Journal of Machine Learning Research: 1
2. IEEE Transactions on Information Theory: 1
3. Annals of Applied Probability
4. Annals of Statistics
5. Stochastic Processes and their Applications
6. Annales de l'IHP : Probabilités et Statistiques: 1
7. ESAIM Probability & Statistics: 1

8. IEEE Transactions on Aerospace and Electronic Systems: 2

9. IEEE Transactions on Signal Processing

Major conferences in the field and number of papers coauthored by members of the project–team that have been accepted during the evaluation period.

1. Winter Simulation Conference

2. International Conference on Information Fusion: 8

3. IEEE Conference on Decision and Control: 1

3.2 Software

No software has been produced during the evaluation period.

3.3 Valorization and technology transfer

Results and expertise in the field of signal processing for digital communications, especially equalization and synchronization, that belong to the former EPI SIGMA2, have been the object of valorization and technology transfer to Alcatel Lucent, with the participation of Frédéric Cérou in a series of ANR projects COHDEQ40 and TCHATER within the Télécommunications programme, and STRADE within the Réseaux du Futur et Services programme. A patent [P1] has been jointly submitted in September 2008 by Alcatel Lucent and INRIA.

Expertise in the field of particle filtering, that truly belongs to ASPI, has been the object of valorization and technology transfer to Thalès Communications, under the DGA project NCT on terrain–aided navigation. A different strategy has been used here, with François Le Gland spending at least one day a week in Thalès Communications premises in Massy, over a 6 months period from September 2008 to March 2009.

3.4 Teaching

As an associate professor, Arnaud Guyader (yearly total, 192 hours) is teaching analysis, statistics and probability at université de Rennes 2, and he is also a member of the committee of *"oraux blancs d'agrégation de mathématiques"* for ENS Cachan at Ker Lann.

François Le Gland (yearly total, 52 hours) gives a course on Bayesian filtering and particle approximation at ENSTA (École Nationale Supérieure de Techniques Avancées, 21 hours), a course on hidden Markov models at Télécom Bretagne (6 hours), a course on linear and nonlinear filtering at ENSAI (École Nationale de la Statistique et de l'Analyse de l'Information, 10 hours) and a course on Kalman filtering and hidden Markov models at université de Rennes 1, master in electronics and telecommunications, track in signal and image processing, embedded systems, and control (15 hours)

The master course at université de Rennes 1 is part of the MATISSE (Mathématiques, Télécommunications, Informatique, Signal, Systèmes, Électronique) doctoral programme.

3.5 Visibility

Arnaud Guyader and Frédéric Cérou have organized, together with Éric Matzner–Løber from université de Rennes 2, the workshop JSTAR (Journées de Statistique de Rennes) held at INRIA Rennes — Bretagne Atlantique in October 2009, `http://www.irisa.fr/aspi/fcerou/STAR2009/Accueil.html`.

4 External Funding

(K € / civil year)	2006	2007	2008	2009	2010
INRIA Research initiatives (all budget goes to coordinator)					
ARC ADOQA					
ARC RARE					
National initiatives					
ANR FIL			21	21	28
ANR PREVASSEMBLE				2	5
ANR NEBBIANO		21	32	32	12
ANR COHDEQ40		57	28		10
ANR TCHATER			11		8
ANR STRADE					17
European projects					
Aeronautics and Space iFLY		25			15
Industrial contracts					
France Télécom RD	30				
Thalès Communications					35
Thalès Communications (CIFRE)			15	15	15
DGA/CTSN					10
Total	30	103	107	70	155
(number / academic year)	05/06	06/07	07/08	08/09	09/10
Grants and scholarships (*)					
ONERA/DTIM PhD					1
ONERA/DPRS PhD				1	1
DGA/CEP PhD		1	1	1	
CIFRE PhD			1	1	1
INRIA Post–Doc.		1			
INRIA Délégation			1	1	
Total	0	2	3	4	3

(∗) other than those supported by one of the above projects

4.1 INRIA Research initiatives (ARC)

ADOQA (from 01/2005 to 12/2006). This ARC was coordinated by EPI CLIME, INRIA Paris — Rocquencourt and CEREA/ENPC. Other partners was EPI IDOPT (now MOISE), INRIA Grenoble — Rhône–Alpes. One objective of ADOQA was to investigate sequential methods (as opposed to variational methods) for data assimilation of intrinsically nonlinear models, i.e. coupling of numerical models and measured data. In principle, a data assimilation algorithm should propagate uncertainties through the probability distribution of the state variables, whereas current sequential algorithms, such as the ensemble Kalman filter (EnKF), only propagate the first two moments. For large–scale systems (physical state of the atmosphere, of the ocean, chemical composition of the atmosphere, etc.), the direct implementation of sequential Monte Carlo methods seems impractical, and simplified, reduced–order models should be used. Our contribution has been to better understand and compare on simple examples such as the three–dimensional Lorenz model, the qualitative behaviour and the performance in terms of the sample size, of the ensemble Kalman filter [O11] and other sequential data assimilation methods, with the qualitative behaviour and the performance of particle filters. The interest of ASPI for these questions started on this occasion, with the internship of Vu–Duc Tran, later continued in her thesis [T3].

RARE (from 01/2006 to 12/2007) This ARC was coordinated by EPI ARMOR (now DIONYSOS), INRIA Rennes — Bretagne Atlantique. The industrial partners were EDF RD and DGAC/DSNA, and the international partners were CWI (The Netherlands) and University of Bamberg (Germany). The objective of RARE was to design and evaluate various Monte Carlo techniques (importance sampling, importance splitting, cross–entropy, etc.), for the simulation of rare but critical events, in several important domain of applications (communication networks, financial risk management, air traffic management, etc.). Our contribution has been to better understand the asymptotic behaviour of importance splitting methods, where intermediate less critical events are introduced, and where trajectories that manage to reach an upper level are replicated into a number of offsprings. Splitting can be achieved in many different ways: in *fixed splitting* for instance, each successful trajectory is given a prescribed deterministic number (possibly depending on the generation number) of offsprings, whereas in *fixed effort* splitting, there is a prescribed deterministic number of trajectories alive at each generation, which amounts to sample with replacement from the successful trajectories at the current stage of the algorithm, and in *fixed performance* splitting a random number of trajectories is simulated, until a prescribed deterministic number of successful trajectories is obtained [B5]. It appears that importance splitting can be interpreted in terms of Feynman–Kac distributions, which makes it possible not only to approximate the probability of the rare but critical event, but also to learn which critical trajectories are responsible for the critical event to occur [J6]. Challenging issues that have been investigated include the automatic selection of the intermediate sets and their number: asymptotic results have been obtained in the one–dimensional case [J9], while in the multi–dimensional case a preliminary objective would be the efficient choice of the importance function, used to define the intermediate sets as level sets. These contributions were collected in a survey article [B6] included in the monograph published on this occasion.

4.2 National initiatives

FIL (ALLOC 2856, from 01/2008 to 12/2010). This ANR project within the Télécommunications programme is coordinated by Thalès Alenia Space. Academic partners are LAAS (laboratoire d'architecture et d'analyse des systèmes), TeSA consortium including ENAC (école nationale de l'aviation civile) and ISAE (institut des sciences de l'aéronautique et de l'espace). Industrial partners are Microtec and Silicom.

The overall objective is to study and demonstrate information fusion algorithms for localisation of pedestrian users in an indoor environment, where GPS solution cannot be used. The design combines

- a pedestrian dead–reckoning (PDR) unit, providing noisy estimates of the linear displacement, angular turn, and possibly of the level change through an additional pression sensor,
- range and / or proximity measurements provided by beacons at fixed and known locations, and possibly indirect distance measurements to access points, through a measure of the power signal attenuation,

- constraints provided by an indoor map of the building (map-matching),
- collaborative localisation when two users meet and exchange their respective position estimates.

Particle methods have been proposed [C2], [C23] as the basic information fusion algorithm for the centralized server–based version, including adaption of the sample size using KLD–sampling [O12], or collaboration between users [O13].

PREVASSEMBLE (ALLOC 3767, from 01/2009 to 12/2011). This ANR project within the COSINUS (conception and simulation) programme, is coordinated by École Normale Supérieure, Paris. The other partner is Météo–France. This is a collaboration with Étienne Mémin (EPI FLUMINANCE, INRIA Rennes — Bretagne Atlantique) and with Anne Cuzol and Valérie Monbet from université de Bretagne Sud in Vannes.

The contribution of ASPI to this project is to continue the comparison [T3], [B4], [R10] of sequential data assimilation methods, such as the ensemble Kalman filter (EnKF) and the weighted ensemble Kalman filter (WEnKF), with particle filters. This comparison will be made on the basis of asymptotic variances, as the ensemble or sample size goes to infinity, and also on the impact of dimension on small sample behaviour.

NEBBIANO (ALLOC 2229, from 01/2007 to 11/2009). Arnaud Guyader was the coordinator of this ANR project, within the SETIN (sécurité et informatique) programme. This is a collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique and Thomson Security Labs during the year 2009) and Pierre Del Moral (EPI ALEA, INRIA Bordeaux — Sud Ouest). The other partner was LIS–INPG in Grenoble. There are mainly two strategic axes in NEBBIANO: watermarking and independent component analysis, and watermarking and rare event simulations. To protect copyright owners, user identifiers are embedded in purchased content such as music or movie. This is basically what we mean by watermarking. This watermarking is to be “invisible” to the standard user, and as difficult to find as possible. When content is found in an illegal place (e.g. a P2P network), the right holders decode the hidden message, find a serial number, and thus they can trace the traitor, i.e. the client who has illegally broadcast their copy. However, the task is not that simple as dishonest users might collude. For security reasons, anti–collusion codes have to be employed. Yet, these solutions (also called weak traceability codes) have a non–zero probability of error defined as the probability of accusing an innocent. This probability should be, of course, extremely low, but it is also a very sensitive parameter: anti–collusion codes get longer (in terms of the number of bits to be hidden in content) as the probability of error decreases. Fingerprint designers have to strike a trade–off, which is hard to conceive when only rough estimation of the probability of error is known. The major issue for fingerprinting algorithms is the fact that embedding large sequences implies also assessing reliability on a huge amount of data which may be practically unachievable without using rare event analysis. Our task within this project is to adapt our methods for estimating rare event probabilities to this framework, and provide watermarking designers with much more accurate false detection probabilities than the bounds currently found in the literature. We have already applied these ideas to some randomized watermarking schemes and obtained much sharper estimates of the probability of accusing an innocent. A patent [P2] has been jointly submitted in August 2008 by INRIA and Bretagne Valorisation.

The next series of ANR projects has been the occasion of a transfer to Alcatel Lucent of results and expertise in the field of signal processing for digital communications, especially equalization and synchronization, that belong to the former EPI SIGMA2.

COHDEQ40 (ALLOC 2205, from 12/2006 to 11/2009). This first ANR project within the Télécommunications programme was coordinated by Alcatel–Lucent. This is a collaboration with Jean–Jacques Fuchs (EPI TEMICS, INRIA Rennes — Bretagne Atlantique). The project COHDEQ40 intends to demonstrate the potential of coherent detection associated with digital signal processing for the next generation high density 40Gb/s WDM systems optimized for transparency and flexibility. Key integrated optoelectronic components and specific algorithms will be developed and system evaluation performed. The INRIA task is to develop these signal processing algorithms needed to recover the message on the decoder side. This makes full use of the knowledge and expertise that belong to the former EPI SIGMA2 on equalization and synchronization techniques involved in digital communications [C14]. A patent [P1] has been jointly submitted in September 2008 by Alcatel Lucent and INRIA.

TCHATER (ALLOC 2801, from 01/2008 to 12/2010). This second ANR project within the Télécommunications programme was coordinated by Alcatel–Lucent. The primary goal of the TCHATER project is to demonstrate a coherent terminal operating at 40Gb/s using *real–time* digital signal processing and efficient polarization division multiplexing. The terminal will benefit to next-generation high information-spectral density optical networks, while offering straightforward compatibility with current 10Gbit/s networks. It will require that advanced high–speed electronic components, especially analog–to–digital converters, are designed within the project. Specific algorithms for polarisation demultiplexing and forward error correction with soft decoding will also have to be developed.

STRADE (ALLOC 4402, from 11/2009 to 10/2012). This third ANR project within the Réseaux du Futur et Services programme was coordinated by Alcatel–Lucent Bell Labs France. The focus of this project is to reduce the impact of nonlinear effect. The objective is twofold: specify, design, realize and evaluate fibres of reduced nonlinear effects by firstly increasing the effective area to unprecedented values and secondly, by splitting optical power along two modes, using bimodal propagation. While the first step is ambitious but primarily relies in the evolution of current fibre technologies, the second is disruptive and requires not only deep changes in fibre technologies but also new advanced transmitter / receiver equipment, preferably based on coherent detection. Naturally, bimodal propagation also brings another key advantage, namely a twofold increase of system capacity.

4.3 European project

iFLY (ALLOC 2399, from 05/2007 to 08/2010). This FP6 project within the Aeronautics and Space programme is coordinated by National Aerospace Laboratory (NLR) (The Netherlands), and can be seen as a follow–up of the previous FP5 project HYBRIDGE within the IST programme. Eighteen academic and industrial partners were involved, and actual collaboration occurred with Politecnico di Milano (Italy), University of Twente (The Netherlands), ETH Zürich (Switzerland), DGAC/DSNA and ENAC (France), and Honeywell (Czech Republic). The objective of iFLY is to develop both an advanced airborne self separation design and a highly automated air traffic management (ATM) design for en–route traffic, which takes advantage of autonomous aircraft operation capabilities and which is aimed to manage a three to six times increase in current en–route traffic levels. The proposed research combines expertise in air transport human factors, safety and economics with analytical and Monte Carlo simulation methodologies. The contribution of ASPI to this project concerns the work package on accident risk assessment methods and their implementation using conditional Monte Carlo methods, especially for large scale stochastic hybrid systems: designing and studying variants [O2] suited for hybrid state space (resampling per mode, marginalization) are currently investigated.

4.4 Associated teams and other international projects

No participation in associated teams or in international (other than european) projects.

4.5 Industrial contracts

Geolocalization (ALLOC 851, from 05/2005 to 08/2006). This project was supported by France Télécom RD (now Orange Labs). The objective was to implement and assess the performance of particle filtering in localisation and tracking of mobile terminals in a wireless network, using network measurements (received signal power strength and possibly TOA (time of arrival)) and a database of reference measurements of the signal power strength, available in a few points or in the form of a digital map (power attenuation map). Generic algorithms have been specialized to the indoor context (wireless local area network, e.g. WiFi) and to the outdoor context (cellular network, e.g. GSM). In particular, constraints and obstacles such as building walls in an indoor environment, street, road or railway networks in an outdoor environment, have been represented in a simplified manner, using a prior model on a graph, e.g. a Voronoï graph as in similar experiments in mobile robotics [O23]. The findings of this work is that localisation in outdoor applications using measurements of the signal power strength alone is not yet operational, whereas the situation is much more favorable in indoor applications. This is because the digital maps available for GSM are obtained by running numerical propagation models that do not capture small scale variations, and the solution would be to use additional measurements, such as TOA (time of arrival).

Terrain–aided navigation (ALLOC 2857, from 09/2007 to 08/2010). This collaboration with Thalès Communications is supported by DGA (Délégation Générale à l'Armement) under the PEA project NCT on terrain–aided navigation, and is related with the supervision of the CIFRE thesis of Nordine El Baraka. The overall objective is to study innovative algorithms for terrain–aided navigation, and to demonstrate these algorithms on four different situations involving different platforms, inertial navigation units, sensors and georeferenced databases. The thesis also considers the special use of image sensors (optical, infra–red, radar, sonar, etc.) for navigation tasks, based on correlation between the observed image sequence and a reference image available on–board in the database. Marginalized particle filters [O32] and regularized particle filters [O24] have been implemented, and several propositions have been studied to adapt the sample size, such as KLD–sampling [O12], which could be useful in the case of a poor initial information, or if the platform flies over a poorly informative area [C15]. Besides particle methods, which are proposed as the basic navigation algorithm, simpler algorithms such as the extended Kalman filter (EKF) or the unscented Kalman filter (UKF) have also been investigated.

Optimization of sensors position and activation (ALLOC 4233, from 04/2009 to 03/2011). This project is supported by CTSN (Centre Technique des Systèmes Navals) a DGA (Délégation Générale à l'Armement) entity. This collaboration with Sébastien Paris, from université Paul Cézanne, is related with the supervision of the PhD thesis of Mathieu Chouchane, and was initiated by Jean–Pierre Le Cadre. The objective of this project is to optimize the position and activation times of a few sensors deployed by a platform 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. Because of the many constraints including timing constraints involved in this optimization problem, a stochastic algorithm is preferred here over a deterministic algorithm. The underlying idea is to replace the problem of maximizing a cost function (the probability of detection) over the possible configurations (admissible position and activation times) by the apparently simpler problem of sampling a population according to a probability distribution depending on a small parameter, which asymptotically concentrates on the set of global maxima of the cost function, as the small parameter goes to zero. The usual approach here is to use the cross–entropy method [O31, O5].

5 Objectives for the next four years

The proposition for the next four years can be divided in three parts. The first part consists in the continuation of ongoing works, some of which are connected with work-packages of the european project iFLY or the ANR projects FIL and PREVASSEMBLE. To this first group belong

- comparison of parametric and nonparametric learning of the optimal (zero variance) importance distribution,
- asymptotic behaviour of rare event simulation methods adapted to hybrid stochastic systems [O19], [O2],
- asymptotic behaviour of KLD–sampling [O12], an adaptive sample size algorithm,
- large sample asymptotics of the ensemble Kalman filter and of its weighted version, introduced by Étienne Mémin (EPI FLUMINANCE, INRIA Rennes — Bretagne Atlantique) and his group.

The second part deals with global optimization, a new direction already mentioned at the end of Section 2.5. The third part can be seen as an opportunity to apply recent results obtained by ASPI in functional classification, and described in Section 2.7 above, and to contribute to the theme web of knowledge and services, which has been identified by INRIA in its strategic plan as one of its priority for the next period. It builds upon the (unsuccessful) proposition MESSY submitted to the ARC programme.

5.1 New objective: Global optimization

An issue closely related to rare event simulation 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 [O31, O5], which relies on learning the optimal importance distribution within a prescribed parametric family. On the other hand, multilevel splitting methods developed in ASPI could provide an alternate nonparametric approach to this problem, and it is worth comparing these with new algorithms proposed recently [O18, O3, O30].

Besides theoretical studies, interest in random optimization algorithms such as the cross–entropy method to solve operations research problems can be seen as a legacy of Jean–Pierre Le Cadre: trajectory optimization for a mobile robot localization with map uncertainties, was addressed in the thesis [T1] of Francis Céleste (DGA/CEP), in a series of conference papers [C3, C5, C6, C7, C8, C9], and in the survey paper [B1], trajectory optimization to minimize the detection of an aircraft is currently addressed in the doctoral project of Renaud Cariou (DGA/CELAR), and optimization of sensors position and activation is the object of a contract with DGA/CTSN and is currently addressed in the doctoral project of Mathieu Chouchane, a PhD student of Sebastien Paris.

5.2 New objective: Statistical analysis of recommendation systems

A collaborative recommendation system is an information filtering technique which recommends certain items to a customer, based on his profile and on data collected from other customers. Most contributions so far have focused on proposing heuristics and practical solutions [O1]. The machine learning community is very active in this respect, but there are still too few statistical studies addressing the behaviour of these many heuristics. High dimensional data and missing data are the two main issues to be addressed.

To fix ideas, consider the example of a movie recommendation system: given a set of $(d+1)$ movies, each of n customers has rated the movies he / she has already seen, which means that missing data cannot be ignored. When a new customer enters in, who has already seen some of the d first movies, the objective is to predict which rating this new customer would give to the $(d+1)$–th movie, so that the system would know whether it should recommend it or not. From the statistical viewpoint, this boils down to a regression problem: estimate the conditional expectation of the variable Y, the rating of the $(d+1)$–th movie by the new customer, given the variables $(X_1, \cdots, X_d)$, the d previous ratings by the same customer. The basic idea is quite simple: the rating by the new customer will be some average of the ratings given to this $(d+1)$–th movie by the subpopulation of the n customers he / she is closest to in some sense. This is clearly a nearest neighbor approach.

For recommendation systems, the qualitative notion of closeness is usually measured in terms of a cosine similarity, and a first consistency result has been obtained in this framework, for the nearest neighbor predictor, by Gérard Biau (université Pierre et Marie Curie, ENS Paris and EPI CLASSIC, INRIA Paris — Rocquencourt), Benoît Cadre (ENS Cachan, Bruz and IRMAR) and Laurent Rouvière (CREST ENSAI and IRMAR). This result relies however on an unrealistic assumption, which ignores missing data, and some extra work is needed to take missing data into account.

Since n and d are very large, even finding the nearest neighbors is already a problem, which is also met in indexing of video documents, and is addressed by Hervé Jégou (EPI TEXMEX, INRIA Rennes — Bretagne Atlantique). Learning algorithms, that have low complexity and are adapted to high dimension, are needed here, and a natural idea is to consider ensemble methods, such as bagging and boosting already studied [J1], in collaboration with Gérard Biau.

Finally, the problem can also be related with fingerprinting: here, the rating of the $(d+1)$–th movie is seen as a combination of the ratings of some other movies in the set, and the question is to find which movies. The key issue is to find a suitable score function to measure the closeness between rating profiles, and to come with a cheap implementation of this function. These ideas are taken from the area of protection of digital documents, and have been studied recently [B2, B3] in collaboration with Teddy Furon (EPI TEMICS, INRIA Rennes — Bretagne Atlantique).

6 Bibliography of the project-team

Doctoral dissertations and "Habilitation" theses

[T1] Francis Céleste. *Cartes Incertaines et Planification Optimale pour la Localisation d'un Engin Autonome*. Thèse de Doctorat, Université de Rennes 1, Rennes, February 2010.

[T2] Adrien Ickowicz. *Estimation de Paramètres Dynamiques à l'Aide d'un Réseau de Capteurs*. Thèse de Doctorat, Université de Rennes 1, Rennes, April 2010.

[T3] Vu-Duc Tran. *Assimilation de Données : les Propriétés Asymptotiques du Filtre de Kalman d'Ensemble*. Thèse de Doctorat, Université de Bretagne Sud, Vannes, June 2009.

Articles in referred journals

[J1] Gérard Biau, Frédéric Cérou, and Arnaud Guyader. On the rate of convergence of the bagged nearest neighbor estimate. *Journal of Machine Learning Research*, 11:687–712, February 2010.

[J2] Gérard Biau, Frédéric Cérou, and Arnaud Guyader. On the rate of convergence of the functional k–nearest neighbor estimates. *IEEE Transactions on Information Theory*, IT–56(4):2034–2040, April 2010.

[J3] Fabien Campillo, Rivo Rakotozafy, and Vivien Rossi. Parallel and interacting Markov chain Monte Carlo algorithm. *Mathematics and Computers in Simulation*, 79(12):3424–3433, August 2009.

[J4] Fabien Campillo and Vivien Rossi. Convolution particle filter for parameter estimation in general state–space models. *IEEE Transactions on Aerospace and Electronic Systems*, 45(3):1063–1072, July 2009.

[J5] Frédéric Cérou, Pierre Del Moral, and Arnaud Guyader. A nonasymptotic variance theorem for unnormalized Feynman–Kac particle models. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 2010. To appear.

[J6] Frédéric Cérou, Pierre Del Moral, François Le Gland, and Pascal Lezaud. Genetic genealogical models in rare event analysis. *ALEA, Latin American Journal of Probability and Mathematical Statistics*, 1:181–203, 2006. Paper 01–08.

[J7] Frédéric Cérou, Teddy Furon, and Arnaud Guyader. Experimental assessment of the reliability for watermarking and fingerprinting schemes. *EURASIP Journal on Information Security*, 2008. Article ID 414962.

[J8] Frédéric Cérou and Arnaud Guyader. Nearest neighbor classification in infinite dimension. *ESAIM : Probability and Statistics*, 10:340–355, 2006.

[J9] Frédéric Cérou and Arnaud Guyader. Adaptive multilevel splitting for rare event analysis. *Stochastic Analysis and Applications*, 25(2):417–443, March 2007.

[J10] Frédéric Cérou and Arnaud Guyader. Adaptive particle techniques and rare event estimation. *ESAIM : Proceedings*, 19:65–72 (electronic), 2007.

[J11] Adrien Ickowicz and Jean-Pierre Le Cadre. On the effect of data contamination on track purity. *IEEE Transactions on Aerospace and Electronic Systems*, 2010. To appear.

[J12] François Le Gland. Combined use of importance weights and resampling weights in sequential Monte Carlo methods. *ESAIM : Proceedings*, 19:85–100 (electronic), 2007.

[J13] François Le Gland and Nadia Oudjane. A sequential algorithm that keeps the particle system alive. In Henk Blom and John Lygeros, editors, *Stochastic Hybrid Systems : Theory and Safety Critical Applications*, number 337 in Lecture Notes in Control and Information Sciences, pages 351–389. Springer–Verlag, Berlin, 2006.

Articles in books and book chapters

[B1] Frédéric Dambreville, Francis Céleste, and Cécile Simonin. Modélisation probabiliste de politiques et leurs optimisations pour la planification de capteurs. In Patrick Siarry, editor, *Optimisation en Traitement du Signal et de l'Image*, Traité Information, Commande, Communication (IC2), chapter 6. Hermès Science, Paris, 2007.

[B2] Teddy Furon, Arnaud Guyader, and Frédéric Cérou. On the design and optimization of Tardos probabilistic fingerprinting codes. In Kaushal Solanki, Kenneth Sullivan, and Upamanyu Madhow, editors, *10th International Workshop on Information Hiding, Santa Barbara 2008*, volume 5284 of *Lecture Notes in Computer Science*, pages 341–356. Springer–Verlag, Berlin, May 2008.

[B3] Teddy Furon, Luis Pérez-Freire, Arnaud Guyader, and Frédéric Cérou. Estimating the minimal length of Tardos codes. In Stefan Katzenbeisser and Ahmad-Reza Sadeghi, editors, *11th International Workshop on Information Hiding, Darmstadt 2009*, number 5806 in Lecture Notes in Computer Science, pages 176–190. Springer, June 2009.

[B4] François Le Gland, Valérie Monbet, and Vu-Duc Tran. Large sample asymptotics for the ensemble Kalman filter. In Dan O. Crisan and Boris L. Rozovskii, editors, *Handbook on Nonlinear Filtering*. Oxford University Press, Oxford, 2010. To appear.

[B5] François Le Gland and Nadia Oudjane. A sequential algorithm that keeps the particle system alive. In Henk Blom and John Lygeros, editors, *Stochastic Hybrid Systems : Theory and Safety Critical Applications*, number 337 in Lecture Notes in Control and Information Sciences, pages 351–389. Springer–Verlag, Berlin, 2006.

[B6] Pierre L'Écuyer, François Le Gland, Pascal Lezaud, and Bruno Tuffin. Splitting methods. In Gerardo Rubino and Bruno Tuffin, editors, *Monte Carlo Methods for Rare Event Analysis*, chapter 3, pages 39–61. John Wiley & Sons, Chichester, 2009.

Publications in conferences and workshops

[C1] Élise Arnaud and François Le Gland. SMC with adaptive resampling : Large sample asymptotics. In *Proceedings of the 2009 Workshop on Statistical Signal Processing, Cardiff 2009*, pages 481–484. IEEE–SPS, September 2009.

[C2] Pierre Blanchart, Liyun He, and François Le Gland. Information fusion for indoor localization. In *Proceedings of the 12th International Conference on Information Fusion, Seattle 2009*, pages 2083–2090. ISIF, July 2009.

[C3] Nicolas Broguière, Francis Céleste, and Frédéric Dambreville. Optimal team decision : a cross–entropic coordination approach. In *Proceedings of the 11th International Conference on Information Fusion, Cologne 2008*, pages 1222–1228. ISIF, July 2008.

[C4] Fabien Campillo and Vivien Rossi. Convolution particle filtering for parameter estimation in general state–space models. In *Proceedings of the 45th IEEE Conference on Decision and Control, San Diego*. IEEE–CSS, December 2006.

[C5] Francis Céleste, Frédéric Dambreville, and Jean-Pierre Le Cadre. Optimal path planning using cross–entropy method. In *Proceedings of the 9th International Conference on Information Fusion, Florence 2006*. ISIF, July 2006.

[C6] Francis Céleste, Frédéric Dambreville, and Jean-Pierre Le Cadre. Optimal strategies for mobile robots based on the cross–entropy algorithm. In *Proceedings of the 18th Workshop on Machine Learning for Signal Processing, Maynooth 2006*. IEEE–SPS, September 2006.

[C7] Francis Céleste, Frédéric Dambreville, and Jean-Pierre Le Cadre. Evaluation of a robot learning and planning via extreme value theory. In *Proceedings of the 10th International Conference on Information Fusion, Québec 2007*. ISIF, July 2007.

[C8] Francis Céleste, Frédéric Dambreville, and Jean-Pierre Le Cadre. Path planning for multiple features based localization. In *Proceedings of the 5th International Conference on Informatics in Control, Automation and Robotics, Funchal 2008*, pages 214–219. IEEE–SMC, May 2008.

[C9] Francis Céleste, Frédéric Dambreville, and Jean-Pierre Le Cadre. Optimized trajectories for mobile robot localization with map uncertainty. In *Proceedings of the 15th Symposium on System Identification (SYSID), Saint Malo 2009*. IFAC, July 2009. Paper WeA6.3.

[C10] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. Rare event simulation for a static distribution. In *Proceedings of the 7th International Workshop on Rare Event Simulation (RESIM), Rennes*, September 2008.

[C11] Frédéric Cérou, Pierre Del Moral, Arnaud Guyader, François Le Gland, Pascal Lezaud, and Hélène Topart. Some recent improvements to importance splitting. In *Proceedings of the 6th International Workshop on Rare Event Simulation (RESIM), Bamberg 2006*, October 2006.

[C12] Frédéric Cérou, Teddy Furon, Arnaud Guyader, and Cyrille Jégourel. Estimating the probability of false alarm for a zero–bit watermarking technique. In *16th International Conference on Digital Signal Processing, Santorin 2009*, July 2009.

[C13] Gabriel Charlet, Jérémie Renaudier, Haik Mardoyan, Oriol Bertran Pardo, Frédéric Cérou, Patrice Tran, and Sébastien Bigo. 12.8 Tbit/s transmission of 160 channels PDM QPSK (160x2x40 Gb/s) with coherent detection over 2,550 km. In *Proceedings of the 33rd European Conference and Exhibition on Optical Communication (ECOC), Berlin*, September 2007. post–deadline paper.

[C14] Gabriel Charlet, M. Salsi, Haik Mardoyan, Patrice Tran, Jérémie Renaudier, Sébastien Bigo, M. Astruc, P. Sillard, L. Provost, and Frédéric Cérou. Transmission of 81 channels at 40Gbit/s over a transpacific–distance erbium–only link, using PDM–BPSK modulation, coherent detection, and a new large effective area fibre. In *Proceedings of the 34th European Conference and Exhibition on Optical Communication (ECOC), Brussels*, September 2008. paper Th.3.E.3.

[C15] Nordine El Baraka and François Le Gland. Bayesian terrain–aided inertial navigation using an airborne laser scanner. In *Proceedings of the 2010 ION International Technical Meeting, San Diego 2010*. ION, January 2010.

[C16] Julien Guillet and François Le Gland. Using noisy georeferenced information sources for navigation and tracking. In *Proceedings of the IEEE Workshop on Nonlinear Statistical Signal Processing (NSSP), Cambridge*. IEEE–SPS, September 2006.

[C17] Adrien Ickowicz and Jean-Pierre Le Cadre. On the effect of data contamination for multitarget tracking. Part I. In *Proceedings of the 10th International Conference on Information Fusion, Québec 2007*. ISIF, July 2007.

[C18] Adrien Ickowicz and Jean-Pierre Le Cadre. On the effect of data contamination for multitarget tracking. Part II. In *Proceedings of the 10th International Conference on Information Fusion, Québec 2007*. ISIF, July 2007.

[C19] Adrien Ickowicz and Jean-Pierre Le Cadre. A new method for target trajectory estimation within a binary sensor network. In *Proceedings of the Workshop on Multi-Camera and Multi–Modal Sensor Fusion : Algorithms and Applications (M2SFA2), Marseilles 2008*, October 2008.

[C20] Adrien Ickowicz and Jean-Pierre Le Cadre. Target tracking within a binary sensor network. In *Proceedings of the International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Sydney 2008*, December 2008.

[C21] Adrien Ickowicz and Jean-Pierre Le Cadre. Target trajectory estimation within a sensor network. In *Proceedings of the 11th International Conference on Information Fusion, Cologne 2008*. ISIF, July 2008.

[C22] Adrien Ickowicz and Jean-Pierre Le Cadre. Bi–target tracking within a binary sensor network. In *Proceedings of the 12th International Conference on Information Fusion, Seattle 2009*. ISIF, July 2009.

[C23] Damien Kubrak, François Le Gland, Liyun He, and Yann Oster. Multi–sensor fusion for localization. Concept and simulation results. In *Proceedings of the 2009 ION Conference on Global Navigation Satellite Systems, Savannah 2009*. ION, September 2009.

Patents

[P1] Frédéric Cérou, Gabriel Charlet, Jérémie Renaudier, and Massimiliano Salsi. A decision directed algorithm for adjusting a polarization demultiplexer in a coherent detection optical receiver, September 2008. European Patent EP–0829902.9.

[P2] Frédéric Cérou, Teddy Furon, and Arnaud Guyader. Validation de schémas de verrous numériques en watermarking et fingerprinting, August 2008. Brevet FR–0804584.

Internal reports (unpublished, submitted, to appear, etc.)

[R1] Gérard Biau, Frédéric Cérou, and Arnaud Guyader. On the rate of convergence of the bagged nearest neighbor estimate. Rapport de Recherche 6860, INRIA, February 2009. http://hal.inria.fr/inria-00363875/en/.

[R2] Gérard Biau, Frédéric Cérou, and Arnaud Guyader. On the rate of convergence of the functional k-NN estimates. Rapport de Recherche 6861, INRIA, February 2009. http://hal.inria.fr/inria-00364555/en/.

[R3] Fabien Campillo and Vivien Rossi. Convolution particle filters for parameter estimation in general state–space models. Rapport de Recherche 5939, INRIA, June 2006. http://hal.inria.fr/docs/00/08/21/61/PDF/RR-5939.pdf.

[R4] Fabien Campillo and Vivien Rossi. Parallel and interacting Markov chain Monte Carlo method. Rapport de Recherche 6008, INRIA, October 2006. http://hal.inria.fr/docs/00/11/08/82/PDF/RR-6008.pdf.

[R5] Frédéric Cérou, Pierre Del Moral, Teddy Furon, and Arnaud Guyader. Rare event simulation for a static distribution. Rapport de Recherche 6792, INRIA, January 2009. http://hal.inria.fr/inria-00363875/en/.

[R6] Frédéric Cérou, Pierre Del Moral, and Arnaud Guyader. A nonasymptotic variance theorem for unnormalized Feynman–Kac particle models. Rapport de Recherche 6716, INRIA, November 2008. http://hal.inria.fr/inria-00337392/fr/.

[R7] Frédéric Cérou, Pierre Del Moral, François Le Gland, and Pascal Lezaud. Genetic genealogical models in rare event analysis. Rapport de Recherche 5878, INRIA, April 2006. ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-5878.pdf.

[R8] Frédéric Cérou, Teddy Furon, and Arnaud Guyader. Experimental assessment of the reliability for watermarking and fingerprinting schemes. Rapport de Recherche 6636, INRIA, September 2008. http://hal.inria.fr/inria-00319768/fr/.

[R9] Arnaud Guyader, Nicolas W. Hengartner, and Éric Matzner-Løber. Iterative Monte Carlo for extreme quantiles and extreme probabilities. Technical report, February 2010. http://www.sites.univ-rennes2.fr/laboratoire-statistique/AGUYADER/doc/ghm.pdf.

[R10] François Le Gland, Valérie Monbet, and Vu-Duc Tran. Large sample asymptotics for the ensemble Kalman filter. Rapport de Recherche 7014, INRIA, August 2009. http://hal.inria.fr/docs/00/40/90/60/PDF/RR-7014.pdf.

[R11] François Le Gland and Nadia Oudjane. A sequential algorithm that keeps the particle system alive. Publication Interne 1783, IRISA, February 2006. `ftp://ftp.irisa.fr/techreports/2006/PI-1783.pdf`.

A Appendix

Additional references

[O1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: a survey of the state–of–the–art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, KDE–17(6):734–749, June 2005.

[O2] Henk A. P. Blom, Bert Bakker, and Jaroslav Krystul. Rare event estimation for a large scale stochastic hybrid system with air traffic application. In Gerardo Rubino and Bruno Tuffin, editors, *Monte Carlo Methods for Rare Event Analysis*, chapter 9, pages 193–214. John Wiley & Sons, Chichester, 2009.

[O3] Zdravko I. Botev and Dirk P. Kroese. An efficient algorithm for rare–event probability estimation, combinatorial optimization, and counting. *Methodology and Computing in Applied Probability*, 10(4):471–505, December 2008.

[O4] Frédéric Cérou, Pierre Del Moral, François Le Gland, and Pascal Lezaud. Limit theorems for the multilevel splitting algorithm in the simulation of rare events. In Michael E. Kuhl, Natalie M. Steiger, Frank B. Armstrong, and Jeffrey A. Joines, editors, *Proceedings of the 2005 Winter Simulation Conference, Orlando 2005*, pages 682–691, December 2005.

[O5] Pieter-Tjerk De Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. A tutorial on the cross–entropy method. *Annals of Operations Research*, 134 (Special issue on the Cross-Entropy Method for Combinatorial Optimization, Rare Event Simulation and Neural Computation)(1):19–67, January 2005.

[O6] Pierre Del Moral. *Feynman–Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Probability and its Applications. Springer–Verlag, New York, 2004.

[O7] Pierre Del Moral and Josselin Garnier. Genealogical particle analysis of rare events. *The Annals of Applied Probability*, 15(4):2496–2534, November 2005.

[O8] Pierre Del Moral and Jean Jacod. Interacting particle filtering with discrete observations. In Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 3, pages 43–75. Springer–Verlag, New York, 2001.

[O9] Pierre Del Moral, Jean Jacod, and Philip Protter. The Monte Carlo method for filtering with discrete–time observations. *Probability Theory and Related Fields*, 120(3):346–368, 2001.

[O10] Pierre Del Moral and Pascal Lezaud. Branching and interacting particle interpretation of rare event probabilities. In Henk Blom and John Lygeros, editors, *Stochastic Hybrid Systems : Theory and Safety Critical Applications*, number 337 in Lecture Notes in Control and Information Sciences, pages 277–323. Springer–Verlag, Berlin, 2006.

[O11] Geir Evensen. *Data Assimilation. The Ensemble Kalman Filter*. Springer–Verlag, Berlin, 2006.

[O12] Dieter Fox. Adapting the sample size in particle filters through KLD–sampling. *International Journal of Robotics Research*, 22(12):985–1003, December 2003.

[O13] Dieter Fox, Wolfram Burgard, Hannes Kruppa, and Sebastian Thrun. A probabilistic approach to collaborative multi–robot localization. *Autonomous Robots*, 8(3 (Special issue on Heterogeneous Multi–Robot Systems)):325–344, June 2000.

[O14] Dieter Fox, Jeffery Hightower, Lin Liao, Dirk Schulz, and Gaetano Borriello. Bayesian filtering for location estimation. *IEEE Pervasive Computing*, 2(3):24–33, July/September 2003.

[O15] Neil J. Gordon, David J. Salmond, and Adrian F. M. Smith. Novel approach to nonlinear / non–Gaussian Bayesian state estimation. *IEE Proceedings, Part F*, 140(2):107–113, April 1993.

[O16] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, and Per-Johan Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, SP–50(2 (Special issue on Monte Carlo Methods for Statistical Signal Processing)):425–437, February 2002.

[O17] Andrei N. Kolmogorov and Vladimir M. Tikhomirov. ε–entropy and ε–capacity of sets in functional spaces. In *Twelve papers on Algebra and Real Functions*, volume 17 of *AMS Translations — Series 2*, pages 277–364. AMS, Providence, 1961.

[O18] Dirk P. Kroese, Sergey Porotsky, and Reuven Y. Rubinstein. The cross–entropy method for continuous multiextremal optimization. *Methodology and Computing in Applied Probability*, 8(3):383–407, September 2006.

[O19] Jaroslav Krystul. *Modelling of Stochastic Hybrid Systems with Applications to Accident Risk Assessment*. Ph.D Thesis, Department of Applied Mathematics, University of Twente, Enschede, September 2006.

[O20] Sanjeev R. Kulkarni and Steven E. Posner. Rates of convergence of nearest neighbor estimation under arbitrary sampling. *IEEE Transactions on Information Theory*, IT–41(4):1028–1039, July 1995.

[O21] François Le Gland and Nadia Oudjane. Stability and uniform approximation of nonlinear filters using the Hilbert metric, and application to particle filters. *The Annals of Applied Probability*, 14(1):144–187, February 2004.

[O22] Pierre L'Ecuyer, Valérie Demers, and Bruno Tuffin. Rare events, splitting, and quasi–Monte Carlo. *ACM Transactions on Modeling and Computer Simulation*, 17(2 (Special issue honoring Perwez Shahabuddin)), April 2007. Article 9.

[O23] Lin Liao, Dieter Fox, Jeffrey Hightower, Henry Kautz, and Dirk Schulz. Voronoi tracking : Location estimation using sparse and noisy sensor data. In *Proceedings of the IEEE / RSJ International Conference on Intelligent Robots and Systems, Las Vegas 2003*, pages 723–728. IEEE / RSJ, October 2003.

[O24] Christian Musso, Nadia Oudjane, and François Le Gland. Improving regularized particle filters. In Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, chapter 12, pages 247–271. Springer–Verlag, New York, 2001.

[O25] Nadia Oudjane. *Stabilité et Approximations Particulaires en Filtrage Non–Linéaire — Application au Pistage.* Thèse de Doctorat, Université de Rennes 1, Rennes, December 2000. `ftp://ftp.irisa.fr/techreports/theses/2000/oudjane.pdf`.

[O26] Stephen B. Pope. *Turbulent Flows.* Cambridge University Press, Cambridge, 2000.

[O27] David Preiss. Gaussian measures and the density theorem. *Commentationes MathematicæUniversitatis Carolinæ*, 22(1):181–193, 1981.

[O28] Branko Ristić, Sanjeev Arulampalam, and Neil J. Gordon. *Beyond the Kalman Filter : Particle Filters for Tracking Applications.* Artech House, Boston, 2004.

[O29] Vivien Rossi. *Filtrage Non–Linéaire par Noyaux de Convolution — Application à un Procédé de Dépollution Biologique.* Thèse de Doctorat, École Nationale Supérieure Agronomique de Montpellier, Montpellier, December 2004.

[O30] Reuven Y. Rubinstein. The Gibbs cloner for combinatorial optimization, counting and sampling. *Methodology and Computing in Applied Probability*, 11(4):491–549, December 2009.

[O31] Reuven Y. Rubinstein and Dirk 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.

[O32] Thomas Schön, Fredrik Gustafsson, and Per-Johan Nordlund. Marginalized particle filters for mixed linear / nonlinear state–space models. *IEEE Transactions on Signal Processing*, SP–53(7):2279–2289, July 2005.

[O33] Charles J. Stone. Consistent nonparametric regression (with discussion). *The Annals of Statistics*, 5(4):595–645, July 1977.

[O34] Gábor Tardos. Optimal probabilistic fingerprint codes. In *Proceedings of the 35th annual ACM symposium on Theory of Computing, San Diego 2003*, pages 116–125. ACM, 2003.

[O35] Sebastian Thrun, Wolfram Burgard, and Dieter Fox. *Probabilistic Robotics.* Intelligent Robotics and Autonomous Agents. The MIT Press, Cambridge, MA, 2005.

B Acronyms

ANR: Agence Nationale de la Recherche

CIFRE: Convention Industrielles de Formation par la Recherche

ARC: Action de Recherche Coopérative INRIA

EPI: équipe–projet INRIA (research group)

DGA: Délégation Générale à l'Armement

- CEP: Centre d'Expertise Parisien (Arcueil)
- CELAR: Centre d'Électronique de l'Armement (Bruz)
- CTSN: Centre Technique des Systèmes Navals (Toulon)
- SPNuM: Service des Programmes Nucléaires et de Missiles

funding programs

- PEA: Programme d'Étude Amont
- REI: Recherche Exploratoire et Innovation

ONERA: Office National d'Études et Recherches Aérospatiales

- DEMR: Département Électromagnétisme et Radar
- DPRS: Département Prospective et Synthèse
- DTIM: Département Traitement de l'Information et Modélisation'

EDF: Électricité de France

- OSIRIS: Optimisation, Simulation, Risque et Statistiques

DGAC: Délégation Générale de l'Aviation Civile

- DSNA: Direction des Services de la Navigation Aérienne

NLR: National Aerospace Laboratory (The Netherlands)

ENAC: École Nationale de l'Aviation Civile

ENSAE: École Nationale Supérieure de l'Administration Économique (Malakoff)

ENSAI: École Nationale de la Statistique et de l'Analyse de l'Information (Bruz)

- CREST: Centre de Recherche en Économie et en Statistique

ENSTA: École Nationale Supérieure de Techniques Avancées (Paris)

IRMAR: Institut de Recherche Mathématique de Rennes, UMR 6625

LSIS: Laboratoire des Sciences de l'Information et des Systèmes, UMR 6168 (Marseille)

CIRAD: Centre de Coopération Internationale en Recherche Agronomique pour le Développement

DSTO: Defence Science and Technology Organization, Australia

NCAR: National Center for Atmospheric Research (Boulder)

NERSC: Nansen Environmental and Remote Sensing Center (Bergen)

LANL: Los Alamos National Laboratory

ION: Institute of Navigation

ISIF: International Society of Information Fusion