



IN PARTNERSHIP WITH:  
**CNRS**

**Université de Lorraine**

Activity Report 2016

# Project-Team **BIGS**

Biology, genetics and statistics

IN COLLABORATION WITH: Institut Elie Cartan de Lorraine (IECL)

RESEARCH CENTER  
**Nancy - Grand Est**

THEME  
**Computational Biology**



## Table of contents

<b>1. Members</b>	<b>2</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Research Program</b>	<b>3</b>
3.1. Introduction	3
3.2. Stochastic modeling	3
3.3. Estimation and control for stochastic processes	3
3.4. Algorithms and estimation for graph data	4
3.5. Regression and machine learning	4
<b>4. Application Domains</b>	<b>5</b>
4.1. Tumor growth	5
4.2. Photodynamic therapy	5
4.3. Genomic data and micro-organisms population study	6
4.4. Epidemiology and e-health	6
4.5. Dynamics of telomeres	6
<b>5. Highlights of the Year</b>	<b>6</b>
<b>6. New Software and Platforms</b>	<b>6</b>
<b>7. New Results</b>	<b>7</b>
7.1. Stochastic modeling	7
7.1.1. Spatial and spatio-temporal modeling	7
7.1.2. Modelisation of response to chemotherapy in gliomas	7
7.1.3. Time-changed extremal process as a random sup measure	7
7.1.4. Fast and Exact synthesis of some operator scaling Gaussian random fields	8
7.1.5. DNA sequences analysis	8
7.2. Estimation and control for stochastic processes	8
7.2.1. Piecewise-deterministic Markov processes	8
7.2.2. Statistics of Markov chains	9
7.2.3. Realtime Tracking of the Photobleaching Trajectory during Photodynamic Therapy	9
7.2.4. Stochastic simulation and design of numerical experiments for the prediction of nanoparticles/X-ray interactions in radiotherapy.	10
7.2.5. Complexity analysis of Policy Iteration	10
7.2.6. Approximate Dynamic Programming for Markov Games	11
7.3. Algorithms and estimation for graph data	11
7.3.1. Modelisation of networks of multiagent systems	11
7.3.2. Compression and analysis of trees	11
7.4. Regression and machine learning	12
7.4.1. Aggregated methods for covariates selection in high-dimensional data under dependence	12
7.4.2. Clustering of the values of a response variable and simultaneous covariate selection using a stepwise algorithm	12
7.4.3. Death or hospitalization scoring for heart failure patients	12
7.4.4. Sequential linear regression with online standardized data	13
7.4.5. Mixed-effects ARX Model Identification of Dynamical Biological Systems	13
7.4.6. Uniform asymptotic certainty bands for the conditional cumulative distribution function	13
7.4.7. Omnibus tests for regression models	13
<b>8. Bilateral Contracts and Grants with Industry</b>	<b>14</b>
8.1.1.1. Transgene 1. (2016-2017)	14
8.1.1.2. Transgene 2. (2016-2017)	14
8.1.1.3. SAFRAN Aircraft Engines (2016-2019)	14
<b>9. Partnerships and Cooperations</b>	<b>14</b>
9.1. National Initiatives	14

---

9.2. European Initiatives	15
<b>10. Dissemination</b> .....	<b>15</b>
10.1. Promoting Scientific Activities	15
10.1.1. Scientific Events Organisation	15
10.1.2. Scientific Events Selection	15
10.1.3. Journal	15
10.1.3.1. Member of the Editorial Boards	15
10.1.3.2. Reviewer - Reviewing Activities	15
10.1.4. Invited Talks	16
10.1.5. Leadership within the Scientific Community	16
10.1.6. Scientific Expertise	16
10.1.7. Research Administration	16
10.2. Teaching - Supervision - Juries	16
10.2.1. Teaching	16
10.2.2. Supervision	17
10.2.3. Juries	18
10.3. Popularization	18
<b>11. Bibliography</b> .....	<b>19</b>

## **Project-Team BIGS**

*Creation of the Team: 2009 January 01, updated into Project-Team: 2011 January 01*

### **Keywords:**

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

- 3. - Data and knowledge
  - 3.1. - Data
    - 3.1.1. - Modeling, representation
  - 3.2. - Knowledge
    - 3.2.3. - Inference
  - 3.3. - Data and knowledge analysis
    - 3.3.1. - On-line analytical processing
    - 3.3.2. - Data mining
    - 3.3.3. - Big data analysis
  - 3.4.1. - Supervised learning
  - 3.4.2. - Unsupervised learning
  - 3.4.4. - Optimization and learning
  - 3.4.7. - Kernel methods
- 3.5. - Social networks
  - 3.5.1. - Analysis of large graphs
- 4.9.1. - Intrusion detection
- 6. - Modeling, simulation and control
  - 6.1. - Mathematical Modeling
    - 6.1.2. - Stochastic Modeling (SPDE, SDE)
  - 6.2. - Scientific Computing, Numerical Analysis & Optimization
    - 6.2.3. - Probabilistic methods
    - 6.2.4. - Statistical methods
  - 6.4. - Automatic control
    - 6.4.2. - Stochastic control

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

- 1. - Life sciences
  - 1.1. - Biology
    - 1.1.2. - Molecular biology
    - 1.1.3. - Cellular biology
    - 1.1.6. - Genomics
    - 1.1.11. - Systems biology
    - 1.1.13. - Plant Biology
  - 1.2. - Ecology
  - 1.4. - Pathologies
- 2.2. - Physiology and diseases
  - 2.2.1. - Cardiovascular and respiratory diseases
  - 2.2.3. - Cancer

- 2.3. - Epidemiology
- 2.4. - Therapies
- 5.5. - Materials
- 9. - Society and Knowledge
- 9.1. - Education

## 1. Members

### Research Scientists

Romain Azaïs [Inria, Researcher]  
Bruno Scherrer [Inria, Researcher, HDR]

### Faculty Members

Anne Gégout Petit [Team leader, Univ. Lorraine, Professor, HDR]  
Thierry Bastogne [Univ. Lorraine, Professor, HDR]  
Sandie Ferrigno [Univ. Lorraine, Associate Professor]  
Sophie Mezieres [Univ. Lorraine, Associate Professor]  
Jean Marie Monnez [Univ. Lorraine, Professor, HDR]  
Aurélien Muller-Gueudin [Univ. Lorraine, Associate Professor]  
Pierre Vallois [Univ. Lorraine, Professor, HDR]  
Yann Petot [Univ. Lorraine, PhD Student, Associate Professor until Sep 2017]

### PhD Students

Lévy Batista [Cybernano, Univ. Lorraine]  
Kévin Duarte [CHU Nancy]  
Florine Greciet [SAFRAN, from Apr 2016, granted by CIFRE]  
Clémence Karmann [Inria, from Oct 2016]  
Pauline Guyot [Hôpital Virtuel de Lorraine, from Nov 2016]

### Post-Doctoral Fellow

Florian Bouguet [Inria, from Sep 2016]

### Visiting Scientists

Céline Lacaux [Univ. Avignon, Professor, HDR]  
Samy Tindel [Purdue University, Professor, HDR]

### Administrative Assistant

Céline Simon [Inria]

### Other

Yaojie Shi [Inria, Internship, from Mar 2016 until Sep 2016]

## 2. Overall Objectives

### 2.1. Overall Objectives

BIGS is a team common to Inria, CNRS and Université de Lorraine, via the Institut Élie Cartan (UMR 7502 CNRS-Inria-UL). Our research is mainly focused on stochastic modeling and statistics for a methodological purpose but also aiming at a better understanding of biological systems. BIGS is composed of applied mathematicians whose research interests mainly concern probability and statistics. More precisely, our attention is directed on (1) stochastic modeling, (2) estimation and control for stochastic processes, (3) algorithms and estimation for graph data and (4) regression and machine learning. The main objective of BIGS is to exploit these skills in applied mathematics to provide a better understanding of some issues arising in life sciences, with a special focus on (1) tumor growth, (2) photodynamic therapy, (3) genomic data and micro-organisms population study, (4) epidemiology and e-health and (5) dynamics of telomeres. Each of these items will be detailed in the sequel.

## 3. Research Program

### 3.1. Introduction

We give here the main lines of our research that belongs to the domains of probability and statistics. For a better understanding, we made the choice to structure them in four items. Although this choice was not arbitrary, the outlines between these items are sometimes fuzzy because each of them deals with modeling and inference and they are all interconnected.

### 3.2. Stochastic modeling

Our aim is to propose relevant stochastic frameworks for the modeling and the understanding of biological systems. The stochastic processes are particularly suitable for this purpose. Among them, Markov chains give a first framework for the modeling of population of cells [105], [69]. Piecewise deterministic processes are non diffusion processes also frequently used in the biological context [53], [68], [61], [56]. Among Markov model, we also developed strong expertise about processes derived from Brownian motion and Stochastic Differential Equations [94], [67], [96]. For instance, knowledge about Brownian or random walk excursions [104], [93] helps to analyse genetic sequences and to develop inference about it. However, nature provides us with many examples of systems such that the observed signal has a given Hölder regularity, which does not correspond to the one we might expect from a system driven by ordinary Brownian motion. This situation is commonly handled by noisy equations driven by Gaussian processes such as fractional Brownian motion or (in higher dimensions of the parameter) fractional fields. The basic aspects of these differential equations are now well understood, mainly thanks to the so-called *rough paths* tools [80], but also invoking the Russo-Vallois integration techniques [95]. The specific issue of Volterra equations driven by fBm, which is central for the subdiffusion within proteins problem, is addressed in [54]. Many generalizations (Gaussian or not) of this model have been recently proposed, see for instance [44] for some Gaussian locally self-similar fields, [73] for some non-Gaussian models, [47] for anisotropic models. Our team has thus contributed [52], [74], [73], [75], [87] and still contributes [46], [48], [47], [76], [64] to this theoretical study: Hölder continuity, fractal dimensions, existence and uniqueness results for differential equations, study of the laws to quote a few examples. On the other hand, because of the observation of longitudinal data for each subject in medicine, we have to care about the random effect due to the subject and to choose adapted models like mixed effect models [77], [42], [43]. In the context of health-care and cost-effectiveness analysis, we are also interested in model of aggregation of different criteria. For this purpose, we develop research about fuzzy binary measures and Choquet integral [63], [81].

### 3.3. Estimation and control for stochastic processes

When one desires to confront theoretical probabilistic models with real data, statistical tools and control of the dynamics are obviously crucial. As matter of course, we develop inference about stochastic processes that we use for modeling, it is the heart of some of our projects. Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy.

The monograph [72] is a good reference on the basic estimation techniques for diffusion processes. Some attention has been paid recently to the estimation of the coefficients of fractional or multifractional Brownian motion according to a set of observations. Let us quote for instance the nice surveys [40], [51]. On the other hand, the inference problem for diffusions driven by a fractional Brownian motion has been in its infancy. A good reference on the question is [103], dealing with some very particular families of equations, which do not cover the cases of interest for us. We also recently proposed least-square estimators for these kind of processes [50], [88]. Inference about PDMP is also a recent subject that we want to develop. Our team has a good expertise about inference of the rate jump and the kernel of PDMP [38], [39], [37], [2]. However, there are many directions to go further into. For instance, previous works made the assumption of a complete observation of jumps and mode, that is unrealistic in practice. We want to tackle the problem of inference of "Hidden PDMP". It could be also interesting to investigate estimation followed by optimal

control for ergodic PDMP. About pharmacokinetics modeling inference, several papers have been reported for the application of system identification techniques. But two issues were ignored in these previous works: presence of timing noise and identification from longitudinal data. In [41], we have proposed a bounded-error estimation algorithm based on interval analysis to solve the parameter estimation problem while taking into consideration uncertainty on observation time instants. Statistical inference from longitudinal data based on mixed effects models [77] can be performed by the *Monolix* software (<http://lixoft.com/products/monolix/>) developed by the Monolix group chaired by Marc Lavielle and France Mentré, and supported by Inria. We used it to estimate tumor growth in [42].

We consider the control of stochastic processes within the framework of Markov Decision Processes [90] and their generalization known as multi-player stochastic games [102], with a particular focus on infinite-horizon problems. In this context, we are interested in the complexity analysis of standard algorithms, as well as the proposition and analysis of numerical approximate schemes for large problems in the spirit of [45]. Regarding complexity, a central topic of research is the analysis of the Policy Iteration algorithm, which has made significant progress in the last years [108], [89], [66], [57], [101], but is still not fully understood. For large problems, we have a long experience of sensitivity analysis of approximate dynamic programming algorithms for Markov Decision Processes [99], [98], [100], [79], [97], and we currently investigate whether/how similar ideas may be adapted to multi-player stochastic games.

### 3.4. Algorithms and estimation for graph data

A graph data structure consists of a set of nodes, together with a set of (either unordered or ordered) pairs of these nodes called edges. This type of data is frequently used in various domains of application (in particular in biology) because they provide a mathematical representation of many concepts such as physical or biological structures and networks of relationship in a population. Some attention has recently been focused in the group on modeling and inference for graph data.

Suppose that we know the value of  $p$  variables on  $n$  subjects (in many applications, we have  $n \ll p$ ). Inference network consists in evaluating the link between two variables knowing the others. [106] gives a very good introduction and many references about network inference and mining. Gaussian Graphical model is a convenient framework to infer network between quantitative variables: there is an edge between two variables if the partial correlation between them is non zero. So the problem is to compute the partial correlations through the concentration matrix. Many methods are available to infer and test partial correlations in the context  $n \ll p$  [106], [82], [60], [62]. However, when dealing with abundance data, because of inflated zero data, data are far from gaussian assumption. Some authors work only with the binary "presence-absence" indicator via log-linear [65]. Models for inflated zero variables are not used for network inference and we want to develop them.

Among graphs, trees play a special role because they offer a good model for many biological concepts, from RNA to phylogenetic trees through plant structures. Our research deals with several aspects of tree data. In particular, we work on statistical inference for this type of data under a given stochastic model (critical Galton-Watson trees for example): in this context, the structure of the tree depends on an integer-valued distribution that we estimate from the observation of either only one tree, or a forest. We also work on lossy compression of trees via linear directed acyclic graphs. These methods make us able to compute distances between tree data faster than from the original structures and with a high accuracy. These results are valuable in the context of very large trees arising for instance in biology of plants.

### 3.5. Regression and machine learning

Regression models or machine learning aim at inferring statistical links between a variable of interest and covariates. It also aims at clustering subjects or variables in set homogeneous sets. In biological study, it is always important to develop adapted learning methods both in the context of "standard" data and also for very massive or online data.



A first approach for regression of quantitative variable is the non-parametric estimation of its cumulative distribution function. Many methods are available to estimate conditional quantiles and test dependencies [86], [70]. Among them we have developed nonparametric estimation through local analysis via polynomial [58], [59] and we want to study properties of this estimator in order to derive measure of risk like confidence band and test. We study also many other regression models like survival analysis, spatio temporal models with covariates. Among the multiple regression models, we want to test, thanks to simulation methods, validity of their assumptions. Tests of this kind are called omnibus test. An omnibus test is an overall test that examines several assumptions together, the most known omnibus test is the one for testing gaussianity (that examines both skewness and kurtosis [55]).

As it concerns the analysis point of high dimensional data, our view on the topic relies on the so-called *French data analysis school*, and more specifically on Factorial Analysis tools. In this context, stochastic approximation is an essential tool (see Lebart's paper [78]), which allows one to approximate eigenvectors in a stepwise manner. A systematic study of Principal Component and Factorial Analysis has then been led by Monnez in the series of papers [85], [83], [84], in which many aspects of convergences of online processes are analyzed thanks to the stochastic approximation techniques. BIGS aims at performing accurate classification or clustering by taking advantage of the possibility of updating the information "online" using stochastic approximation algorithms [71]. We focus on several incremental procedures for regression and data analysis like linear and logistic regressions and PCA. We also focus the biological context of high-throughput bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. The inference of the modeling conclusions from a sample of wells to the whole population requires to account for the inter-individual variability within the modeling procedure. One solution consists in using mixed effects models but up to now no similar approach exists in the field of dynamical system identification. As a consequence, we aim at developing a new solution based on an ARX (Auto Regressive model with eXternal inputs) model structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameters.

## 4. Application Domains

### 4.1. Tumor growth

Cancer is the result of inter-dependent multi-scale phenomena and this is mainly why the understanding of its spread is still an unsolved problem. In integrative biology, mathematical models play a central role; they help biologists and clinicians to answer complex questions through numerical simulations and statistical analyses. The main issue here is to better understand and describe the role of cell damage heterogeneity and associated mutant cell phenotypes in the therapeutic responses of cancer cell populations submitted to a radiotherapy sessions during *in vitro* experiments. The cell heterogeneity is often described as randomness in mathematical modeling and different representations, such as Markov chains, branching processes and even stochastic differential equations, have been recently used.

### 4.2. Photodynamic therapy

Since 1988, some control system scientists and biologists at the CRAN<sup>1</sup> have worked together to develop the photodynamic therapy (PDT in the sequel), an alternative treatment for cancer, by means of a model-based approach. The global aim in this direction is to use statistical as well as mechanistic models in order to improve the response reproducibility, help biologists and chemists in the design of new photosensitizing agents and provide insight into complex phenomena associated with oncogenesis, tumor angiogenesis and interactions with the treatment. This heavily relies on the production of accurate and simple enough models involving various type of stochastic processes, such as Markov chains, branching processes and stochastic differential equations. The main questions here concern generally identification or estimation properties, but simulation issues can be important too.

<sup>1</sup>Centre de Recherche en Automatique de Nancy, [http://www.cran.uhp-nancy.fr/francais/themes\\_rech/sbs/beam/index.php](http://www.cran.uhp-nancy.fr/francais/themes_rech/sbs/beam/index.php)

### 4.3. Genomic data and micro-organisms population study

Generation genomic technologies allow clinicians and biomedical researchers to drastically increase the amount of genomic data collected on large cohort of patients and populations. We want to contribute to a better understanding of the correlations between gene through their expression data, of the structure of ARN and of the genetic bases of drug response and disease and to detect significant sequences characterizing a gene. For instance the biopharmaceutical company Transgene has recently contacted us to analyse their genomic and proteomic data particularly for the purpose to find markers of the success of therapies that they develop against cancer.

Network inference has also applications for the analysis of micro-organisms population, that we apply to micro-organism inside and around the truffle through a collaboration with INRA Nancy. We want also study other specific complex microbial communities like that found at tree roots in order to characterize phenotype of the tree. There is also application in human health (for instance identification of network between bacteria inside colon).

### 4.4. Epidemiology and e-health

Trough J.-M. Monnez and his collaborator Pr E. Albuissou, BIGS is stakeholder of projects with University Hospital of Nancy that is FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARterial AGEing; leader: Pr Athanase BENETOS), RHU Fight HF (Fighting Heart Failure; leader: Pr Patrick ROSSIGNOL), and "Handle your heart", team responsible for the creation of a drug prescription support software for the treatment of heart failure. All these projects are in the context of personalized medicine and deal with biomarkers research; prognostic value of quantitative variables and events and scoring of heart failure. Other collaborations with clinicians concern foetopathology and cancer again.

### 4.5. Dynamics of telomeres

A telomere is a region of repetitive and non coding nucleotide sequences at each end of a chromosome. The telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Mathematical modeling of telomere dynamics is recent [36], [107], [91], [49]. Trough a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, and some members of Inria team TOSCA, we want to work in three connected directions: (1) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior of the distribution of lengths, quasi-stationarity, etc); (2) use these properties to develop new statistical methods for estimating the various parameters; and (3) find and use a suitable methodology for the analysis of the available data (Pr Benetos) for instance for the study of the length distribution for a subject and its evolution.

## 5. Highlights of the Year

### 5.1. Highlights of the Year

BIGS team has organised a two-days workshop "Rencontres des équipes Inria travaillant sur le cancer" that took place Paris in March. 10 inria teams were present. The program is available on <https://team.inria.fr/bigs/workshopcancer/>.

## 6. New Software and Platforms

### 6.1. AGH

KEYWORDS: statistical analysis, ordered trees

#### SCIENTIFIC DESCRIPTION

The Matlab toolbox AGH provides methods for statistical analysis of ordered trees from their Harris paths in a user-friendly environment. More precisely it allows to easily compute estimators of the relative scale of trees which share the same shape. These estimators have been introduced for Galton-Watson trees conditioned on their number of nodes but may be computed for any ordered tree. The theoretical study of these estimators is presented in the associated paper [30] which should be consulted in parallel.

#### FUNCTIONAL DESCRIPTION

The Matlab toolbox AGH provides methods for statistical analysis of ordered trees from their Harris paths in a user-friendly environment.

- Participants: Romain Azaïs, Alexandre Genadot, Benoît Henry
- Contact: romain.azais@inria.fr
- URL: <http://agh.gforge.inria.fr>

## 7. New Results

### 7.1. Stochastic modeling

#### 7.1.1. *Spatial and spatio-temporal modeling*

Participants: A. Gégout-Petit

External collaborators: Y. Cao, S. Li, L. Guerin-Dubrana (Inra Bordeaux)

In the framework of a collaboration with INRA Bordeaux about the esca-illness of vines, Anne Gégout-Petit with Shuxian Li developed different spatial models and spatio-temporal models for different purposes: (1) study the distribution and the dynamics of esca vines in order to tackle the aggregation and the potential spread of the illness (2) propose a spatio-temporal model in order to capture the dynamics of cases and measure the effects of environmental covariates. For purpose (1), we propose different test based on the join count statistics, a paper is accepted for publication [5]. We also developed a two-step centered autologistic model for the study of the dynamic of propagation. This work has been presented as invited paper in [20] and is in preparation for publication.

#### 7.1.2. *Modelisation of response to chemotherapy in gliomas*

Participant: S. Wantz-Mézières

External collaborator: J.-M. Moureaux, Y. Gaudeau, M. Ben Abdallah, M. Ouqamra (CRAN, Université de Lorraine), L. Taillandier, M. Blonski (CHU Nancy)

The collaboration with neurologists (CHU Nancy) and automaticians (CRAN) has carried on this year and led to the PhD presentation of M. Ben Abdallah, on December 12, 2016 [17], [16]. We completed the modeling approach by a data analysis one. In the framework of a master 2 project, supervised and non supervised methods have confirmed our results on our local data base. This encourages us to continue our work in extending the data base via a collaboration with Montpellier CHU. Our perspectives are to validate multi-factor models, including biological and anatomopathological factors, and to design a decision-aid tool for praticians.

#### 7.1.3. *Time-changed extremal process as a random sup measure*

Participant: Céline Lacaux

External collaborator: Gennady Samorodnitsky

In extreme value theory, one of the major topics is the study of the limiting behavior of the partial maxima of a stationary sequence. When this sequence is i.i.d., the unique limiting process is well-known and called the extremal process. Considering a long memory stable sequence, the limiting process is obtained as a simple power time change extremal process. Céline Lacaux and Gennady Samorodnitsky have proved that this limiting process can also be interpreted as a restriction of a self-affine random sup measure. In addition, they have established that this random measure arises as a limit of the partial maxima of the same long memory stable sequence, but in a different space. Their results open the way to propose new self-similar processes with stationary max-increments.

#### 7.1.4. *Fast and Exact synthesis of some operator scaling Gaussian random fields*

Participant: Céline Lacaux

External collaborator: Hermine Biermé

Operator scaling Gaussian random fields, as anisotropic generalizations of self-similar fields, know an increasing interest in the literature. Up to now, such models were only defined through stochastic integrals, without knowing explicitly their covariance functions. In link with this misunderstanding, one of the drawbacks is that no exact method of simulation has been proposed. In view to fill this lack, Hermine Biermé and Céline Lacaux have recently exhibit explicit covariance functions, as anisotropic generalizations of fractional Brownian fields ones. This allows them to propose a fast and exact method to synthetise an operator scaling Gaussian random fields with such a covariance function. Their algorithm is based on the famous circulant embedding matrix method. This is a first piece of work to popularized operator scaling Gaussian random field in anisotropic spatial data modeling.

#### 7.1.5. *DNA sequences analysis*

Participants: P. Vallois

External collaborators: A. Lagnoux and S. Mercier (Toulouse)

In an article accepted at Bioinformatics, the goal is to illustrate different results on the local score distribution assuming an i.i.d. model, especially the one based on the pair (local score,length) and the one on the local score position. We measure with statistical tests how different approximations of the local score distribution fit to simulated sequences. In particular, our simulations show that the popular Karlin & Altschul approximation is not accurate in a wide range of situations. We add to the local score the length of the segment that realises it and we study the induced changes with numerical simulations. We also study specificity and sensitivity for the different methods. We introduce a new one dimensional statistic which is a function of  $H_n^*$  and  $L_n^*$  and we test its distribution. Finally, we estimate the probability that  $H_n^* = H_n$  in different settings.

## 7.2. Estimation and control for stochastic processes

### 7.2.1. *Piecewise-deterministic Markov processes*

Participants: Romain Azaïs, Florian Bouguet, Anne Gégout-Petit, Florine Greciet, Aurélie Muller-Gueudin

External participants: Michel Benaïm (Université de Neuchâtel), Bertrand Cloez (Inra-SupAgro MISTEA), Alexandre Genadot (Inria CQFD, Université de Bordeaux)

A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. This class of stochastic processes offers a wide range of applications, especially in biology (kinetic dietary exposure model and growth of bacteria for example). BIGS' members mainly work on statistical inference techniques for these stochastic processes [2], [29], which is an essential step to build relevant application models. We also investigate the probabilistic properties of these processes [32], [31] as well as the application in reliability to crack growth in some alloy in the industrial context of the PhD thesis of Florine Greciet with SAFRAN Aircraft Engines [33].

In a preprint recently accepted for publication in Electronic Journal of Statistics [2], we focus on the nonparametric estimation problem of the jump rate for piecewise-deterministic Markov processes observed within a long time interval under an ergodicity condition. More precisely, we introduce an uncountable class (indexed by the deterministic flow) of recursive kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. In addition, we propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated. We also discuss the choice of the bandwidth parameters by cross-validation methods. In [29], we state a new characterization of the jump rate when the transition kernel only charges a discrete subset of the state space. We deduce from this result a competitive nonparametric technique for estimating this feature of interest. We state the uniform convergence in probability of the estimator. Both the methodologies have been illustrated on numerical examples and real data.

The article [32] deals with a class of conservative growth-fragmentation equations with a deterministic viewpoint. With the help of Foster-Lyapunov criteria, we study the long-time behavior of some associated piecewise-deterministic Markov process, which represents a typical individual following the dynamics of the equation. If the growth and the fragmentation are balanced, it is possible to provide existence and unicity for the stationary distribution on the process, as well as precise bounds for its tails of distributions in the neighborhoods of both 0 and  $+\infty$ . Our probabilistic results are systematically compared to estimates already obtained with deterministic methods.

In [31], we are interested by the long-time behavior of inhomogeneous-time Markov chains. We put forward an original and unified approach to relate some of their asymptotic properties (stationary distribution, speed of convergence, ...) to the ones of an auxiliary homogeneous-time Markov process. Such results are close to traditional functional limit theorems, but our method differs from the standard “Tightness/Identification” argument; it is based on the notion of asymptotic pseudotrajectories on the space of probability measures. We recover classical results, such as normalized bandit algorithms converging to a piecewise-deterministic Markov process, or weighted random walks or decreasing step Euler schemes approximated with solutions of stochastic differential equations.

### 7.2.2. *Statistics of Markov chains*

Participant: Romain Azaïs

External participants: Bernard Delyon (Université Rennes 1), François Portier (Télécom ParisTech)

Suppose that a mobile sensor describes a Markovian trajectory in the ambient space. At each time the sensor measures an attribute of interest, e.g., the temperature. Using only the location history of the sensor and the associated measurements, the aim of the paper [27] is to estimate the average value of the attribute over the space. In contrast to classical probabilistic integration methods, e.g., Monte Carlo, the proposed approach does not require any knowledge on the distribution of the sensor trajectory. Probabilistic bounds on the convergence rates of the estimator are established. These rates are better than the traditional “root  $n$ ”-rate, where  $n$  is the sample size, attached to other probabilistic integration methods. For finite sample sizes, the good behaviour of the procedure is demonstrated through simulations and an application to the evaluation of the average temperature of oceans is considered.

### 7.2.3. *Realtime Tracking of the Photobleaching Trajectory during Photodynamic Therapy*

Participant: T. Bastogne

Photodynamic therapy (PDT) is an alternative treatment for cancer that involves the administration of a photosensitizing agent, which is activated by light at a specific wavelength. This illumination causes after a sequence of photoreactions, the production of reactive oxygen species responsible for the death of the tumor cells but also the degradation of the photosensitizing agent, which then loose the fluorescence properties. The phenomenon is commonly known as photobleaching process and can be considered as a therapy efficiency indicator. In [8], we present the design and validation of a real time controller able to track a preset photobleaching trajectory by modulating the light impulses width during the treatment sessions. This innovative solution was validated by in vivo experiments that have shown a significantly improvement of reproducibility of the inter-individual photobleaching kinetic. We believe that this approach could lead to personalized photodynamic therapy modalities in the near future.

#### 7.2.4. Stochastic simulation and design of numerical experiments for the prediction of nanoparticles/X-ray interactions in radiotherapy.

Participant: T. Bastogne

The increase of computational environments dedicated to the simulation of nanoparticles (NP)-X-Rays interactions has opened new perspectives in computer-aided-design of nanostructured materials for biomedical applications. Several published studies have shown a crucial need of standardization of these numerical simulations [92]. That is why, we proposed to perform a robustness multivariate analysis in [8]. A gold nanoparticle (GNP) of 100 nm diameter was selected as a standard nano-system activated by a X-ray source placed just below the NP. Two response variables were examined: the dose enhancement in seven different spatial regions of interest around the NP and the duration of the experiments. 9 factors were pre-identified as potentially critical. A Plackett-Burman design of numerical experiments was applied to estimate and test the effects of each simulation factors on the examined responses. Four factors: the working volume, the spatial resolution, the spatial cutoff and the computational mode (parallelization) do not significantly affect the dose deposition results and none except the last one may reduce the computational duration. The energy cutoff may cause significant variations of the dose enhancement in some specific regions of interest: the higher the cutoff, the closer the secondary particles will stop from the GNP. By contrast, the Auger effect as well as the choice of the physical medium and the fluence level clearly appear as critical simulation parameters. Consequently, these four factors may be compulsory examined before comparing and interpreting any simulation results coming from different simulation sessions.

In [9], we address the prediction issue of organometallic nanoparticles (NPs)-based radiosensitization enhancement. The goal was to carry out computational experiments to quickly identify efficient nanostructures and then to preferentially select the most promising ones for the subsequent in vivo studies. To this aim, this interdisciplinary article introduces a new theoretical Monte Carlo computational ranking method and tests it using 3 different organometallic NPs in terms of size and composition. While the ranking predicted in a classical theoretical scenario did not fit the reference results at all, in contrast, we showed for the first time how our accelerated in silico virtual screening method, based on basic in vitro experimental data (which takes into account the NPs cell biodistribution), was able to predict a relevant ranking in accordance with in vitro clonogenic efficiency. This corroborates the pertinence of such a prior ranking method that could speed up the preclinical development of NPs in radiation therapy.

This in-silico approach was tested in [25] to screen radiosensitizing nanoparticles and the results have been validated by in vitro assays.

#### 7.2.5. Complexity analysis of Policy Iteration

Participant: Bruno Scherrer

Given a Markov Decision Process (MDP) with  $n$  states and a total number  $m$  of actions, we study in [10] the number of iterations needed by Policy Iteration (PI) algorithms to converge to the optimal  $\gamma$ -discounted policy. We consider two variations of PI: Howard's PI that changes the actions in all states with a positive advantage, and Simplex-PI that only changes the action in the state with maximal advantage. We show that Howard's PI terminates after at most  $O\left(\frac{m}{1-\gamma} \log\left(\frac{1}{1-\gamma}\right)\right)$  iterations, improving by a factor  $O(\log n)$  a result by Hansen et al., while Simplex-PI terminates after at most  $O\left(\frac{nm}{1-\gamma} \log\left(\frac{1}{1-\gamma}\right)\right)$  iterations, improving by a factor  $O(\log n)$  a result by Ye. Under some structural properties of the MDP, we then consider bounds that are independent of the discount factor  $\gamma$ : quantities of interest are bounds  $\tau_t$  and  $\tau_r$ —uniform on all states and policies—respectively on the *expected time spent in transient states* and *the inverse of the frequency of visits in recurrent states* given that the process starts from the uniform distribution. Indeed, we show that Simplex-PI terminates after at most  $\tilde{O}(n^3 m^2 \tau_t \tau_r)$  iterations. This extends a recent result for deterministic MDPs by Post & Ye, in which  $\tau_t \leq 1$  and  $\tau_r \leq n$ ; in particular it shows that Simplex-PI is strongly polynomial for a much larger class of MDPs. We explain why similar results seem hard to derive for Howard's PI. Finally, under the additional (restrictive) assumption that the state space is partitioned in two sets, respectively states that are transient and recurrent for all policies, we show that both Howard's PI and Simplex-PI terminate after at most  $\tilde{O}(m(n^2 \tau_t + n \tau_r))$  iterations.



### 7.2.6. *Approximate Dynamic Programming for Markov Games*

Participant: Bruno Scherrer

We have made two contributions to the analysis of Approximate Dynamic Programming algorithms for Markov Games.

First, we extend in [21] several non-stationary Reinforcement Learning (RL) algorithms and their theoretical guarantees to the case of discounted zero-sum Markov Games (MGs). As in the case of Markov Decision Processes (MDPs), non-stationary algorithms are shown to exhibit better performance bounds compared to their stationary counterparts. The obtained bounds are generically composed of three terms: 1) a dependency over gamma (discount factor), 2) a concentrability coefficient and 3) a propagation error term. This error, depending on the algorithm, can be caused by a regression step, a policy evaluation step or a best-response evaluation step. As a second contribution, we empirically demonstrate, on generic MGs (called Garnets), that non-stationary algorithms outperform their stationary counterparts. In addition, it is shown that their performance mostly depends on the nature of the propagation error. Indeed, algorithms where the error is due to the evaluation of a best-response are penalized (even if they exhibit better concentrability coefficients and dependencies on gamma) compared to those suffering from a regression error.

Furthermore, we report in [22] theoretical and empirical investigations on the use of quasi-Newton methods to minimize the Optimal Bellman Residual (OBR) of zero-sum two-player Markov Games. First, it reveals that state-of-the-art algorithms can be derived by the direct application of Newton's method to different norms of the OBR. More precisely, when applied to the norm of the OBR, Newton's method results in the Bellman Residual Minimization Policy Iteration (BRMPI) and, when applied to the norm of the Projected OBR (POBR), it results into the standard Least Squares Policy Iteration (LSPI) algorithm. Consequently, new algorithms are proposed, making use of quasi-Newton methods to minimize the OBR and the POBR so as to take benefit of enhanced empirical performances at low cost. Indeed, using a quasi-Newton method approach introduces slight modifications in term of coding of LSPI and BRMPI but improves significantly both the stability and the performance of those algorithms. These phenomena are illustrated on an experiment conducted on artificially constructed games called Garnets.

## 7.3. Algorithms and estimation for graph data

### 7.3.1. *Modelisation of networks of multiagent systems*

Participants: Aurélie Muller-Gueudin

We relate here a collaboration with researchers in Automatic in Nancy (CRAN).

We consider here networks, modeled as a graph with nodes and edges representing the agents and their interconnections, respectively. The connectivity of the network, persistence of links and interactions reciprocity influence the convergence speed towards a consensus.

The problem of consensus or synchronization is motivated by different applications as communication networks, power and transport grids, decentralized computing networks, and social or biological networks.

We then consider networks of interconnected dynamical systems, called agents, that are partitioned into several clusters. Most of the agents can only update their state in a continuous way using only inner-cluster agent states. On top of this, few agents also have the peculiarity to rarely update their states in a discrete way by resetting it using states from agents outside their clusters. In social networks, the opinion of each individual evolves by taking into account the opinions of the members belonging to its community. Nevertheless, one or several individuals can change its opinion by interacting with individuals outside its community. These inter-cluster interactions can be seen as resets of the opinions. This leads us to a network dynamics that is expressed in term of reset systems. We suppose that the reset instants arrive stochastically following a Poisson renewal process.

We have an accepted paper in the journal IEEE Transactions on Automatic Control [6].

### 7.3.2. *Compression and analysis of trees*

Participant: Romain Azaïs

External participants: Jean-Baptiste Durand (ENSIMAG, Inria MISTIS), Christophe Godin (Inria Virtual Plants), Benoît Henry (Inria TOSCA puis Madynes), Alexandre Genadot (Université de Bordeaux, Inria CQFD)

Tree-structured data naturally appear in various fields, particularly in biology where plants and blood vessels may be described by trees, but also in computer science because XML documents form a tree structure. Among trees, the class of self-nested trees presents remarkable compression properties because of the systematic repetition of subtrees in their structure. In a recent preprint [28], we provide a better combinatorial characterization of this specific family of trees. We show that self-nested trees may be considered as a good approximation class of unordered trees. In addition, we compare our approximation algorithms with a competitive approach of the literature on a simulated dataset. On the other hand, the paper [30] is devoted to the estimation of the relative scale of ordered trees that share the same layout. The theoretical study is achieved for the stochastic model of conditioned Galton-Watson trees. New estimators are introduced and their consistency is stated. A comparison is made with an existing approach of the literature. A simulation study shows the good behavior of our procedure on finite-sample sizes. An application to the analysis of revisions of Wikipedia articles is also considered through real data.

## 7.4. Regression and machine learning

### 7.4.1. *Aggregated methods for covariates selection in high-dimensional data under dependence*

Participants: A. Gégout-Petit, A. Muller-Gueudin, Y. Shi

External collaborators: B. Bastien (Transgene, Strasbourg)

In the purpose to select factors linked to the efficiency of a treatment in the context of high dimension (about 100.000 covariates), we have developed a new methodology to select and rank covariates associated to a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology imbricates successively rough selection, clustering of variables, decorrelation of variables using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking through bootstrap replications. Simulations study shows the interest of the decorrelation inside the different clusters of covariates. The methodology was applied to select covariates among genomics, proteomics covariates linked to the success of a immunotherapy treatment for the lung cancer. A paper on the subject is in preparation.

### 7.4.2. *Clustering of the values of a response variable and simultaneous covariate selection using a stepwise algorithm*

Participant: J.-M. Monnez

External collaborator: O. Collignon (LIH Luxembourg)

In supervised learning the number of values of a response variable to predict can be very high. Grouping these values in a few clusters can be useful to perform accurate supervised classification analyses. On the other hand selecting relevant covariates is a crucial step to build robust and efficient prediction models. We propose in this paper an algorithm that simultaneously groups the values of a response variable into a limited number of clusters and selects stepwise the best covariates that discriminate this clustering. These objectives are achieved by alternate optimization of a user-defined selection criterion. This process extends a former version of the algorithm to a more general framework. Moreover possible further developments are discussed in detail [3].

### 7.4.3. *Death or hospitalization scoring for heart failure patients*

Participant: J.-M. Monnez, K. Duarte

External collaborator: E. Albuissou (CHU, Nancy)

The purpose of this study was to define a short term event (death or hospitalization) score for heart failure patients based on the observation of biological, clinical and medical historical variables. Some of them were transformed or winsorized. Two methods of statistical learning were performed, logistic regression and linear discriminant analysis, different variable selection methods were used, on bootstrap samples. Aggregation of classifiers and out-of-bag validation were used. Finally a score taking values between 0 and 100 was established and an odds-ratio was defined in order to support medical decision (writing in progress).



#### **7.4.4. Sequential linear regression with online standardized data**

Participant: J.-M. Monnez, K. Duarte

External collaborator: E. Albuissou

We consider the problem of sequential least square multidimensional linear regression using a stochastic approximation process. The choice of the stepsize may be crucial in this type of process. In order to avoid the risk of numerical explosion which can be encountered, we define three processes with a variable or a constant stepsize and establish their convergence. Finally these processes are compared to classic processes on 11 datasets, 6 with a continuous output and 5 with a binary output, for a fixed total number of observations used and then for a fixed processing time. It appears that the third-defined process with a very simple choice of the stepsize gives usually the best results (paper to be submitted).

#### **7.4.5. Mixed-effects ARX Model Identification of Dynamical Biological Systems**

Participants: T. Bastogne, L. Batista

System identification is a data-driven modeling approach more and more used in biology and biomedicine [26]. In this application context, each assay is always repeated to estimate the response variability. The inference of the modeling conclusions to the whole population requires to account for the inter-individual variability within the modeling procedure. One solution consists in using mixed effects models but up to now no similar approach exists in the field of dynamical system identification. In [23], we propose a new solution based on an ARX (Auto Regressive model with eXternal inputs) structure using the EM (Expectation-Maximisation) algorithm for the estimation of the model parameters. Simulations show the relevance of this solution compared with a classical procedure of system identification repeated for each subject.

In [24], we propose a solution to firstly estimate the Fisher information matrix using the Louis' method and secondly to determine the parameters confidence intervals of an ARX model structure. We show relevance of the proposed solution in simulation and using real in-vitro data coming from realtime cell impedance measurements.

In parallel, we applied the mixed-effect modeling approach to the analysis in vivo responses in order to identify prognostic biomarkers of tumor regrowth after photodynamic therapy [11]. This application corroborated the practical relevance of our model-based approach.

#### **7.4.6. Uniform asymptotic certainty bands for the conditional cumulative distribution function**

Participants: S.Ferrigno, A. Muller-Gueudin, M. Maumy-Bertrand (IRMA, Strasbourg)

In this work, we study the conditional cumulative distribution function and a nonparametric estimator associated to this function. The conditional cumulative distribution function has the advantages of completely characterizing the law of the random considered variable, allowing to obtain the regression function, the density function, the moments and the conditional quantile function. As a nonparametric estimator of this function, we focus on local polynomial techniques described in Fan and Gijbels [ref]. In particular, we use the local linear estimation of the conditional cumulative distribution function.

The objective of this work is to establish uniform asymptotic certainty bands for the conditional cumulative distribution function. To this aim, we give exact rate of strong uniform consistency for the local linear estimator of this function (writing in progress).

#### **7.4.7. Omnibus tests for regression models**

Participants: R.Azaïs, S.Ferrigno, M-J Martinez Marcoux (LJK, Grenoble)

The aim of this collaboration begins is to compare, through simulations, several methods to test the validity of a regression model. These tests can be "directional" in that they are designed to detect departures from mainly one given assumption of the model (for example the regression function, the variance or the error) or global (for example the conditional distribution function). The establishment of such statistical tests require the use of nonparametric estimators various functions (regression, variance, cumulative distribution function). The idea would then be able to build a tool (package R) that allows a user to test the validity of the model it uses through different methods and varying parameters associated with modeling. This work is currently in progress.

## 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

#### 8.1.1. Bilateral Contracts with Industry

##### 8.1.1.1. *Transgene 1. (2016-2017)*

Participants: A. Gégout-Petit, A. Muller-Gueudin, Y. Shi

Transgene (Euronext: TNG), part of Institut Mérieux, is a publicly traded French biopharmaceutical company focused on discovering and developing targeted immunotherapies for the treatment of cancer and infectious diseases. B. Bastien, head of the biostatistics team appeals to BIGS to select covariates among genomics, proteomics expressions linked to the success of a treatment of the lung cancer. This subject was the purpose of the master thesis of Y. Shi and a paper on the subject is in preparation.

##### 8.1.1.2. *Transgene 2. (2016-2017)*

Participants: T. Bastogne, L. Batista, P. Vallois

Transgene (Euronext: TNG), part of Institut Mérieux, is a publicly traded French biopharmaceutical company focused on discovering and developing targeted immunotherapies for the treatment of cancer and infectious diseases. B. Bastien, head of the biostatistics team appeals to BIGS to model data collected in vivo for growth tumor and to measure the effect of the treatment on the dynamics of the tumor.

##### 8.1.1.3. *SAFRAN Aircraft Engines (2016-2019)*

Participants: R. Azaïs, A. Gégout-Petit, F. Greciet

SAFRAN Aircraft Engines designs and products Aircraft Engines. For the design of pieces, they have to understand mechanism of crack propagation under different conditions. It appeals to BIGS for modeling crack propagation with Piecewise Deterministic Markov Processes (PDMP). It is the subject of F. Greciet PhD, granted by ANRT. F. Greciet presented her work during a Fédération Charles Hermite Journey on November the 23th. She was laureat of "Mathématiques, oxygene du monde numérique" poster challenge [33].

## 9. Partnerships and Cooperations

### 9.1. National Initiatives

- *PEPS AMIES* (2016), Apprentissage supervisé pour l'aide au diagnostic, Collaboration Institut Elie Cartan avec la StartUp SD Innovation Frouard. Participants: A. Gégout-Petit, P. Vallois
- *Popart* (2016-2017) In the framework of collaboration with A. Deveau of Inra Nancy, A. Gégout-Petit and A. Muller-Gueudin are included in the Inra "Microbial Ecosystems & Metaomics, Call 2016" Project "Popart" for "Regulation of the Poplar microbiome by its host: is the immune system involved ? ". The aim is to develop methodology for the inference of regulation network between micro-organisms around Poplar. The specificity of the data is the inflation of zeros that has to be taken into account.
- *Intérêt des antiangiogènes dans la potentialisation des thérapies par rayonnement dans le cas des glioblastomes* (2016). Funding organism: Ligue contre le Cancer (CCIR-GE). Leader: N. Thomas (CRAN, U. Lorraine). Participants : C. Lacaux and A. Muller-Gueudin
- (2014-16), A library of Near-Infrared absorbing photosensitizers: tailoring and assessing photo-physical and synergetic photodynamic properties, Funding organism: PHC Bosphore - Campus France, Leader: M. Barberi-Heyob (CRAN), Thierry Bastogne
- GDR 3475 Analyse Multifractale, Funding organism: CNRS, Leader: S. Jaffard (Université Paris-Est), Céline Lacaux

- GDR 3477 Géométrie stochastique, Funding organism: CNRS, Leader: P. Calka (Université Rouen), Céline Lacaux
- FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARterial AGEing ; leader : Pr Athanase BENETOS), Jean-Marie Monnez
- RHU Fight HF (Fighting Heart Failure ; leader : Pr Patrick ROSSIGNOL), located at the University Hospital of Nancy, Jean-Marie Monnez
- Project "Handle your heart", team responsible for the creation of a drug prescription support software for the treatment of heart failure, head: Jean-Marie Monnez

## 9.2. European Initiatives

### 9.2.1. Collaborations in European Programs, Except FP7 & H2020

- Photobrain project. AGuIX theranostic nanoparticles for vascular-targeted interstitial photodynamic therapy of brain tumors, project **EuroNanoMed II**, resp.: M. Barberi-Heyob, (2015-2017), participant: T. Bastogne.
- NanoBit Project. Nanoscintillator-Porphyrin Complexes for Bimodal RadioPhotoDynamic Therapy, project **EuroNanoMed II**, resp.: P. Juzenas, (2016-2018), participant: T. Bastogne.

## 10. Dissemination

### 10.1. Promoting Scientific Activities

#### 10.1.1. Scientific Events Organisation

- BIGS team has organised a two-days workshop "Rencontres des équipes Inria travaillant sur le cancer" in Paris. 10 inria teams were present. <https://team.inria.fr/bigs/workshopcancer/>
- A. Gégout-Petit co-organised the "Health Session" of the day Fédération Charles Hermite- Enterprises, Nancy, January 2016.
- Céline Lacaux participated to the organisation of the following events:
  - European Study Group with Industry, 117th edition, May 2016, Avignon.
  - Session *Statistics* of the 14th Colloque Franco-Roumain de Mathématiques Appliquées, August 2016, Iasi.
  - Conference of GDR 3475 Analyse Multifractale, September 2016, Avignon.

#### 10.1.2. Scientific Events Selection

##### 10.1.2.1. Chair of Conference Program Committees

- A. Gégout-Petit is chair of 2017 "Congrès Francophone International de l'Enseignement de la Statistique" (CFIES), Grenoble, September, 2017.

#### 10.1.3. Journal

##### 10.1.3.1. Member of the Editorial Boards

P. Vallois is in the editorial board of "Risk and Decision Analysis".

##### 10.1.3.2. Reviewer - Reviewing Activities

All the BIGS members are regular reviewers for journals in probability, statistics and machine learning as: Bernoulli, Scandinavian Journal of statistics, Stochastics, Journal of Statistical Planning Inference, Journal of theoretical Biology, IEEE Trans. Biomedical Eng., Theoretical Biology and Medical Modelling, Royal Society of Chemistry, Signal Processing: Image Communication, Mathematical Biosciences, LIDA, Annals of Applied Probability, Annals of Operations Research and Journal of Machine Learning Research, as well as conferences such as ICML, World IFAC Congress, FOSBE, ALCOSP...

#### 10.1.4. Invited Talks

- Anne Gégout-Petit was invited in SADA'2016 conference in Cotonou. [20]
- Bruno Scherrer was an invited speaker in EWRL'2016 workshop in Barcelona.
- Thierry Bastogne was invited the 19 October 2015 by Pr. Luc Leyns at the Vrije Universities Brussels for a talk on the *Statistical Analysis of Cell Impedance Signals for the Characterization of Anti-Cancer Drug Effects*.
- Romain Azaïs was an invited speaker in SSIAB 2016 in Rennes.
- S. Ferrigno : *Nouvelles approches d'estimation de la croissance en Foetopathologie*. Journée "Nouvelle approche de la croissance foetale", Maternité Régionale du CHRU de Nancy (Sept 2016).

#### 10.1.5. Leadership within the Scientific Community

- Anne Gégout-Petit is member of the board of the European Regional Council of the Bernoulli society
- Céline Lacaux is responsible of the *Statistic team*, Laboratory of Mathematic of Avignon (since September 2016)

#### 10.1.6. Scientific Expertise

- T. Bastogne: scientific expert in Biostatistics and Signal Processing in Nanomedicine for CYBER-nano (start-up).

#### 10.1.7. Research Administration

- A. Gégout-Petit: elected member of the laboratory of mathematics "Institut Elie Cartan de Lorraine".
- Céline Lacaux is
  - member of the bord of the SMAI-MAS group,
  - elected member of the council of the Laboratory Mathematics of Avignon,
  - correspondant AMIES pour Avignon,
  - Member of the scientific committee of GDR 3477 Stochastic Geometry.

## 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching

- A. Gégout-Petit : Head of the Master 2 "Ingénierie Mathématique et Outils Informatiques (Mathematical Engineering and Computer Tools)", Université de Lorraine
- A. Gégout-Petit created and is now in charge of cursus CMI in applied mathematics for Lorraine University
- P. Vallois is head of the "Parcours Mathématiques Financières" of the master "Applied mathematics" of Université de Lorraine
- P. Vallois is head of the convention between "Université de Lorraine and Université Hammam Sousse" about master organization. Master ISC (Ingénierie de Systèmes Complexes)
- T. Bastogne is in charge of the spécialité Systèmes & TIC du master Ingénierie de Systèmes Complexes
- T. Bastogne created and is now in charge of professional master: CIIBLE (Cybernétique, Instrumentation, Image en Biologie et medecinE) en M2 with Medicine Faculty of Université de Lorraine
- T. Bastogne created and is now in charge research master << Biosanté Numérique >> with engineering school "Telecom Nancy"
- Master: S. Ferrigno, Experimental designs, 4.5h, M1, fourth year of EEIGM, Université de Lorraine, France

- Master: S. Ferrigno, Data analyzing and mining, 63h, M2, third year of Ecole des Mines, Université de Lorraine, France
- Master: S. Ferrigno, Modeling and forecasting, 43h, M1, second year of Ecole des Mines, Université de Lorraine, France
- Master: S. Ferrigno, Training projects, 18h, M1/M2, second and third year of Ecole des Mines, Université de Lorraine, France
- Master: A. Muller-Gueudin, Probability and Statistics, 160h, second year of ENSEM and ENSAIA, University of Lorraine, France.
- Master: A. Muller-Gueudin, Scientific calculation with Matlab, 20h, second year of ENSAIA, University of Lorraine, France.
- Master: R. Azaïs, Machine learning, 20h, M2, Université de Lorraine and third year of Telecom Nancy, France.
- Master: R. Azaïs, Machine learning, 20h, M1, second year of Ecole des Mines, Université de Lorraine, France.
- Master: J.-M. Monnez, Multivariate Analysis, Master 2 IFM (Ingénierie de la Finance de Marché), until June 2016.
- Master: A. Gégout-Petit, Statistics, modeling, 15h, future teacher, Université de Lorraine, France
- Master: A. Gégout-Petit, Statistics, modeling, data analysis, 80h, master in applied mathematics, Université de Lorraine, France
- Licence: S. Wantz-Mézières, Applied mathematics for management, financial mathematics, Probability and Statistics, 160h, I.U.T. (L1/L2/L3)
- Licence: S. Wantz-Mézières, Probability, 100h, first year in Telecom Nancy engineering school (initial and apprenticeship cursus)
- Master: J.-M. Monnez, Data Analysis, Statistical Learning, Master 2 IMOI (Ingénierie Mathématique et Outils Informatiques), until June 2016.
- Licence: A. Muller-Gueudin, Statistics, 60h, first year of ENSAIA, University of Lorraine, France.
- Licence: S. Ferrigno, Descriptive and inferential statistics, 60h, L2, second year of EEIGM, Université de Lorraine, France
- Licence: S. Ferrigno, Statistical modeling, 60h, L2, second year of EEIGM, Université de Lorraine, France
- Licence: S. Ferrigno, Mathematical and computational tools, 20h, L3, third year of EEIGM, Université de Lorraine, France
- Licence: S. Ferrigno, Training projects, 20h, L1/L3, first, second and third year of EEIGM, Université de Lorraine, France
- Licence: C. Lacaux: Probability and Statistic, 75h, L3, University of Avignon.
- Licence: C. Lacaux: Numerical simulation in probability, 36.75h, L3, University of Avignon.
- Licence: C. Lacaux: Probability and Statistics, 22.5h, L1, University of Avignon.
- Licence: C. Lacaux: Statistic techniques applied to SVT, 25.5h, L3, University of Avignon.
- Licence: C. Lacaux: Statistics, 24h, L2, University of Avignon.

### 10.2.2. Supervision

- PhD : Clémence Karmann, " Network inference for zero-inflated models", Grant : Inria-Cordis. Advisors: A. Gégout-Petit, A. Muller-Gueudin.
- PhD : Florine Gréciet, " Modèles markoviens déterministes par morceaux cachés pour la propagation de fissures", grant CIFRE SAFRAN AIRCRAFT ENGINES, Advisors : R. Azaïs, A. Gégout-Petit.

- PhD : Kévin Duarte, "Aide à la décision médicale et télémédecine dans le suivi de l'insuffisance cardiaque", Advisors : J.-M. Monnez and E. Albuissou.
- PhD : P. Retif. Modeling, digital simulation and analysis of nanoparticles-X ray interaction. Applications to augmented radiotherapy. Theses, Université de Lorraine, Mar. 2016.
- Post-doc: Florian Bouguet. Advisors: Romain Azaïs, Anne Gégout-Petit, Aurélie Muller-Gueudin.
- Post-doc: Benoît Henry (starting in Dec. 2016). Advisors: Romain Azaïs with Inria team Madynes.
- Master: Yaojie Shi, Toulouse School of economics, 2016. « Analyse de données transcriptomiques et protéomiques en oncologie », Advisors: A. Gégout-Petit, A. Muller-Gueudin, B. Bastien (Société Transgene, Strasbourg).
- Master: Yuyan Cao, Toulouse School of economics, 2016. " Spatio-temporal Bayesian models for the analysis of esca disease", with Inra Bordeaux. Advisor: A. Gégout-Petit, L. Guérin-Dubrana.
- Master: Félicie Bonte, Master Ecologie Ecologie Lille, AgroParistech Nancy et Museum National d'Histoire Naturelle 2016. « Etude des modifications de la croissance et du développement des plantes herbacées en forêt en réponse aux changements globaux », co-direction AgroParistech Nancy. Advisors: A. Gégout-Petit, Jean-Claude Gégout, Serge Muller.
- Master: all BIGS members regularly supervise project and internship of master IMOJ students
- Engineering school: all BIGS members regularly supervise project of "Ecole des Mines ", ENSEM or EEIGM students

### 10.2.3. Juries

- HDR, Bruno Scherrer, "Contributions algorithmiques au contrôle optimal stochastique à temps discret et horizon infini", Université de Lorraine, July, 2016, Examiner, A. Gégout-Petit
- HDR, Corine Hahn, "Penser la question didactique pour la formation en alternance dans l'enseignement supérieur. Dispositifs frontières, Statistique et Management." Université Lyon 2, May, 2016, Examiner, A. Gégout-Petit
- PhD, Florian Bouguet, "Etude quantitative de processus de Markov déterministes par morceaux issus de la modélisation", Université de Rennes, June, 2016, Referee: A. Gégout-Petit
- PhD, Houda Ghamlouch, "Modélisation de la dégradation, maintenance conditionnelle et pronostic : usage des processus de diffusion", Université Technologique de Troyes, June, 2016, Referee: A. Gégout-Petit
- PhD, Etienne Baratchart, "Etude quantitative des aspects dynamiques et spatiaux du développement métastatique à l'aide de modèles mathématiques", Université de Bordeaux, February, 2016, Referee: A. Gégout-Petit
- PhD, Johann Cuenin, Sur les modèles Tweedie multivariés, Université de Besancon, December, Examiner: A. Gégout-Petit
- PhD : Clémence Chamard-Jovenin, Impact d'une surexpression d'ER $\alpha$ 36 et/ou d'une exposition aux alkylphénols sur la physiopathologie de la glande mammaire, Université de Lorraine, 9 décembre 2016. Examiner : A. Muller-Gueudin.
- PhD : M. Ben Abdallah, Université de Lorraine, "Un modèle de l'évolution des gliomes diffus de bas grade sous chimiothérapie", December, 12, 2016, jury member: S. Wantz-Mézières.
- PhD : Marc Bourotte, Générateur stochastique de temps multisite basé sur un champ gaussien multivarié, INRA, Équipe BioSP, July, 4th, 2016, President : C. Lacaux.
- PhD : Nhu Dang, Estimation des indices de stabilité et d'autosimilarité par variations de puissances négatives, Laboratoire Jean Kuntzmann Grenoble, July, 5th, 2016, Examiner : C. Lacaux.

## 10.3. Popularization

- A. Gégout-Petit is involved in the promotion of study in the fields of mathematics in Lorraine university. She was very active in the realisation of the video for the promotion of Mathematical studies <http://videos.univ-lorraine.fr/index.php?act=view&id=3236>.
- A. Gégout-Petit participates to the "Table ronde Bourse aux technologies, Big data et industrie du futur", Ecole des Mines de Nancy, November, 2016.
- Animation d'ateliers MATH.en.JEANS en collège dans la région de Nancy (Romain Azaïs, Clémence Karmann)
- S. Ferrigno: Advisor of a group of students, "La main à la Pâte" project, elementary schools, Nancy, January-June 2016
- S. Ferrigno: Advisor of a group of students, "La main à la Pâte" project, Institut médico-éducatif (IME), Commercy, September-December 2016
- S. Ferrigno: Advisor of a group of students, "De Léonard de Vinci au Drone" project, Collège Paul Verlaine, Malzéville, December 2016-February 2017.

## 11. Bibliography

### Publications of the year

#### Doctoral Dissertations and Habilitation Theses

- [1] B. SCHERRER. *Contributions algorithmiques au contrôle optimal stochastique à temps discret et horizon infini*, Université de Lorraine (Nancy), June 2016, Habilitation à diriger des recherches, <https://hal.inria.fr/tel-01400208>

#### Articles in International Peer-Reviewed Journals

- [2] R. AZAÏS, A. MULLER-GUEUDIN. *Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes*, in "Electronic journal of statistics", 2016, <https://hal.archives-ouvertes.fr/hal-01168651>
- [3] O. COLLIGNON, J.-M. MONNEZ. *Clustering of the values of a response variable and simultaneous covariate selection using a stepwise algorithm*, in "Applied Mathematics", 2016, vol. 07, pp. 1639 - 1648 [DOI : 10.4236/AM.2016.715141], <https://hal.inria.fr/hal-01395535>
- [4] C. LACAUX, G. SAMORODNITSKY. *Time-changed Extremal Process as a Random Sup Measure*, in "Bernoulli journal", 2016, vol. 22, n<sup>o</sup> 4, pp. 1979–2000, <https://hal.archives-ouvertes.fr/hal-01102343>
- [5] S. LI, F. BONNEU, J. J. CHADOEUF, D. PICART, A. GÉGOUT-PETIT, L. GUERIN-DUBRANA. *Spatial and Temporal Pattern Analyses 1 of Esca Grapevine Disease in Vineyards in France*, in "Phytopathology", 2017, <https://hal.inria.fr/hal-01205332>
- [6] I.-C. MORARESCU, S. MARTIN, A. GIRARD, A. MULLER-GUEUDIN. *Coordination in networks of linear impulsive agents*, in "IEEE Transactions on Automatic Control", September 2016, vol. 61, n<sup>o</sup> 9, pp. 2402-2415 [DOI : 10.1109/TAC.2015.2492058], <https://hal.archives-ouvertes.fr/hal-01096071>
- [7] G. NICHIL, P. VALLOIS. *Provisioning against borrowers default risk*, in "Insurance: Mathematics and Economics", 2016, vol. 66, pp. 29-43 [DOI : 10.1016/j.insmatheco.2015.10.004], <https://hal.archives-ouvertes.fr/hal-01224520>



- [8] P. RETIF, T. BASTOGNE, M. BARBERI-HEYOB. *Robustness analysis of a Geant4-Gate simulator for nano-radiosensitizers characterization*, in "IEEE Transactions on NanoBioscience", April 2016, vol. 15, n<sup>o</sup> 3, pp. 209-217 [DOI : 10.1109/TNB.2016.2527720], <https://hal.archives-ouvertes.fr/hal-01326803>
- [9] P. RETIF, A. REINHARD, H. PAQUOT, V. JOUAN-HUREAUX, A. CHATEAU, L. SANCEY, M. BARBERI-HEYOB, S. PINEL, T. BASTOGNE. *Monte Carlo simulations guided by imaging to predict the in vitro ranking of radiosensitizing nanoparticles*, in "International Journal of Nanomedicine", November 2016, vol. 11, pp. 6169-6179 [DOI : 10.2147/IJN.S111320], <https://hal.archives-ouvertes.fr/hal-01407060>
- [10] B. SCHERRER. *Improved and Generalized Upper Bounds on the Complexity of Policy Iteration*, in "Mathematics of Operations Research", February 2016, Markov decision processes ; Dynamic Programming ; Analysis of Algorithms [DOI : 10.1287/MOOR.2015.0753], <https://hal.inria.fr/hal-00829532>
- [11] M. TOUSSAINT, S. PINEL, F. AUGER, N. DURIEUX, M. THOMASSIN, E. THOMAS, A. MOUSSARON, D. MENG, F. PLÉNAT, M. AMOUROUX, T. BASTOGNE, C. FROCHOT, O. TILLEMENT, F. LUX, M. BARBERI-HEYOB. *Proton MR spectroscopy and diffusion MR imaging monitoring to predict tumor response to interstitial photodynamic therapy for glioblastoma*, in "Theranostics", 2017, <https://hal.archives-ouvertes.fr/hal-01399256>
- [12] J.-B. TYLCZ, T. BASTOGNE, A. BOURGUIGNON, C. FROCHOT, M. BARBERI-HEYOB. *Realtime tracking of the photobleaching trajectory during photodynamic therapy*, in "IEEE Transactions on Biomedical Engineering", 2017, 8 p. [DOI : 10.1109/TBME.2016.2620239], <https://hal.archives-ouvertes.fr/hal-01205739>

### Invited Conferences

- [13] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX. *Evaluation statistique de la segmentation manuelle de données IRM de gliomes diffus de bas grade*, in "18e Colloque Compression et Représentation des Signaux Audiovisuels, CORESA 2016", Nancy, France, May 2016, <https://hal.archives-ouvertes.fr/hal-01316788>
- [14] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX. *Modèles prédictifs pour les gliomes diffus de bas grade sous chimiothérapie*, in "18e Colloque Compression et Représentation des Signaux Audiovisuels, CORESA 2016", Nancy, France, May 2016, <https://hal.archives-ouvertes.fr/hal-01316799>
- [15] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX. *Modèles prédictifs pour les gliomes diffus de bas grade sous chimiothérapie*, in "Modélisation biostatistique et biomathématique des données d'imagerie en cancérologie", Bordeaux, France, Cancéropôle Grand Sud-Ouest, April 2016, Présentation Poster, <https://hal.archives-ouvertes.fr/hal-01321281>
- [16] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX. *Predictive models for diffuse low-grade glioma patients under chemotherapy*, in "38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'16", Orlando, Florida, United States, August 2016, <https://hal.archives-ouvertes.fr/hal-01316865>
- [17] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX. *Statistical evaluation of manual segmentation of a diffuse low-grade glioma MRI dataset*, in "38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'16", Orlando, Florida, United States, August 2016, <https://hal.archives-ouvertes.fr/hal-01316879>



### International Conferences with Proceedings

- [18] M. ACHARD, S. ACHERAR, P. ALTHUSER, J.-C. ANDRE, A. PHILIPPE, M. BARBERI-HEYOB, F. BAROS, T. BASTOGNE, C. BONINSEGNA, C. BOURA, L. COLOMBEAU, C. FROCHOT, V. JOUAN-HUREAUX, S. GORIA, J. LANDON, A. MOHD GAZZALI, S. PINEL, T. ROQUES-CARMES, N. THOMAS, M. TOUSSAINT, R. VANDERESSE, Z. YOUSSEF. *PDTeam's project: targeting to improve PDT selectivity*, in "Photodynamic Therapy and Photodiagnosis update", Nancy, France, October 2016, pp. Poster PC-028, Présentation Poster, <https://hal.archives-ouvertes.fr/hal-01408912>
- [19] K. DUARTE, J.-M. MONNEZ, E. ALBUISSON. *Valeur pronostique d'une estimation du volume plasmatique dans l'insuffisance cardiaque*, in "48èmes Journées de Statistique de la Société Française de Statistique", Montpellier, France, May 2016, <https://hal.archives-ouvertes.fr/hal-01321101>
- [20] A. GÉGOUT-PETIT, S. LI. *Two-step centered spatio-temporal auto-logistic regression model*, in "SADA416, Applied Statistics for Development in Africa", Cotonou, Benin, November 2016, <https://hal.inria.fr/hal-01394868>
- [21] J. PEROLAT, B. PIOT, B. SCHERRER, O. PIETQUIN. *On the Use of Non-Stationary Strategies for Solving Two-Player Zero-Sum Markov Games*, in "19th International Conference on Artificial Intelligence and Statistics (AISTATS 2016)", Cadiz, Spain, Proceedings of the International Conference on Artificial Intelligences and Statistics, May 2016, <https://hal.inria.fr/hal-01291495>
- [22] J. PÉROLAT, B. PIOT, M. GEIST, B. SCHERRER, O. PIETQUIN. *Softened Approximate Policy Iteration for Markov Games*, in "ICML 2016 - 33rd International Conference on Machine Learning", New York City, United States, June 2016, <https://hal.inria.fr/hal-01393328>

### Conferences without Proceedings

- [23] L. BATISTA, T. BASTOGNE, E.-H. DJERMOUNE. *Identification of dynamical biological systems based on mixed-effect models*, in "31st ACM Symposium on Applied Computing", Pisa, Italy, April 2016, <https://hal.archives-ouvertes.fr/hal-01205738>
- [24] L. BATISTA, T. BASTOGNE, E.-H. DJERMOUNE. *Mixed-effects ARX model identification of dynamical systems*, in "25th Meeting of the Population Approach Group in Europe, PAGE 2016", Lisboa, Portugal, June 2016, PAGE 25 (2016) Abstr 5807 p. , Présentation Poster, <https://hal.archives-ouvertes.fr/hal-01320594>
- [25] P. HÉNA, A. REINHARD, P. RETIF, A. CHATEAU, V. JOUAN-HUREAUX, L. SANCEY, T. BASTOGNE, P. QUETIN, S. PINEL, M. BARBERI-HEYOB. *Screening process of radiosensitizing nanoparticles based on basic in vitro experimental data and in silico analysis*, in "13e Colloque Nano-hybrides", Porquerolles, France, ILM, Lyon 1, May 2016, Présentation Poster, <https://hal.archives-ouvertes.fr/hal-01312945>

### Scientific Books (or Scientific Book chapters)

- [26] M. SCHLIEMANN-BULLINGER, D. FEY, T. BASTOGNE, R. FINDEISEN, P. SCHEURICH, E. BULLINGER. *The experimental side of parameter estimation*, in "Uncertainty in Biology - A Computational Modeling Approach", L. GERIS, D. GOMEZ-CABRERO (editors), Studies in Mechanobiology, Tissue Engineering and Biomaterials, Springer, 2016, vol. 17, pp. 127-154, <https://hal.archives-ouvertes.fr/hal-01256462>

### Other Publications

- [27] R. AZAÏS, B. DELYON, F. PORTIER. *Integral estimation based on Markovian design*, September 2016, 45 pages, <https://hal.archives-ouvertes.fr/hal-01360647>
- [28] R. AZAÏS, J.-B. DURAND, C. GODIN. *Approximation of trees by self-nested trees*, September 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01294013>
- [29] R. AZAÏS, A. GENADOT. *A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions*, June 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01334847>
- [30] R. AZAÏS, A. GENADOT, B. HENRY. *Inference for conditioned Galton-Watson trees from their Harris path*, September 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01360650>
- [31] M. BENAÏM, F. BOUGUET, B. CLOEZ. *Ergodicity of inhomogeneous Markov chains through asymptotic pseudotrajectories*, November 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01401981>
- [32] F. BOUGUET. *A Probabilistic Look at Growth-Fragmentation Equations*, September 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01362555>
- [33] F. GRECIET, R. AZAÏS, A. GEGOUT-PETIT, M. BIRET. *Modèle markovien déterministe par morceaux caché pour la propagation de fissure*, October 2016, Mathématiques, oxygène du numérique, Poster, <https://hal.archives-ouvertes.fr/hal-01398288>
- [34] P. HÉNA, A. REINHARD, P. RETIF, A. CHATEAU, V. JOUAN-HUREAUX, L. SANCEY, T. BASTOGNE, M. BARBERI-HEYOB, S. PINEL. *Pouvoir radiopotentialisant in vitro de nanoparticules hybrides à base de métaux*, November 2016, pp. 59- poster 14, 10ème Forum du Cancéropôle Grand-Est, Poster, <https://hal.archives-ouvertes.fr/hal-01409010>
- [35] A. LAGNOUX, S. MERCIER, P. VALLOIS. *Statistical significance for sequence analysis. Illustration of new results on length and position of the local score*, April 2016, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01301246>

## References in notes

- [36] O. ARINO, M. KIMMEL, G. F. WEBB. *Mathematical modeling of the loss of telomere sequences*, in "Journal of theoretical biology", 1995, vol. 177, n<sup>o</sup> 1, pp. 45–57
- [37] R. AZAÏS. *A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process*, in "ESAIM: Probability and Statistics", 2014, vol. 18, pp. 726–749
- [38] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes*, in "Annales de l'Institut Henri Poincaré, Probabilités et Statistiques", Institut Henri Poincaré, 2013, vol. 49, n<sup>o</sup> 4, pp. 1204–1231
- [39] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Nonparametric estimation of the jump rate for piecewise-deterministic Markov processes*, in "Scandinavian Journal of Statistics", 2014, 26 p.

- [40] J. M. BARDET, G. LANG, G. OPPENHEIM, A. PHILIPPE, S. STOEV, M. TAQQU. *Semi-parametric estimation of the long-range dependence parameter: a survey*, in "Theory and applications of long-range dependence", Birkhauser Boston, 2003, pp. 557-577
- [41] T. BASTOGNE, S. MÉZIÈRES-WANTZ, N. RAMDANI, P. VALLOIS, M. BARBERI-HEYOB. *Identification of pharmacokinetics models in the presence of timing noise*, in "Eur. J. Control", 2008, vol. 14, n<sup>o</sup> 2, pp. 149–157, <http://dx.doi.org/10.3166/ejc.14.149-157>
- [42] T. BASTOGNE, A. SAMSON, P. VALLOIS, S. WANTZ-MEZIERES, S. PINEL, D. BECHET, M. BARBERI-HEYOB. *Phenomenological modeling of tumor diameter growth based on a mixed effects model*, in "Journal of theoretical biology", 2010, vol. 262, n<sup>o</sup> 3, pp. 544–552
- [43] L. BATISTA, T. BASTOGNE, E.-H. DJERMOUNE. *Identification of dynamical biological systems based on random effects models*, in "37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2015", Milan, Italy, August 2015, <https://hal.archives-ouvertes.fr/hal-01159193>
- [44] A. BENASSI, S. JAFFARD, D. ROUX. *Elliptic Gaussian random processes*, in "Rev. Mat. Iberoamericana", 1997, vol. 13, n<sup>o</sup> 1, pp. 19–90
- [45] D. BERTSEKAS, J. TSITSIKLIS. *Neurodynamic Programming*, Athena Scientific, 1996
- [46] H. BIERMÉ, C. LACAUX. *Hölder regularity for operator scaling stable random fields*, in "Stochastic Process. Appl.", 2009, vol. 119, n<sup>o</sup> 7, pp. 2222–2248, <http://dx.doi.org/10.1016/j.spa.2008.10.008>
- [47] H. BIERMÉ, C. LACAUX, H.-P. SCHEFFLER. *Multi-operator Scaling Random Fields*, in "Stochastic Processes and their Applications", 2011, vol. 121, n<sup>o</sup> 11, pp. 2642-2677, MAP5 2011-01 [DOI : 10.1016/J.SPA.2011.07.002], <http://hal.archives-ouvertes.fr/hal-00551707/en/>
- [48] H. BIERMÉ, C. LACAUX, Y. XIAO. *Hitting probabilities and the Hausdorff dimension of the inverse images of anisotropic Gaussian random fields*, in "Bull. Lond. Math. Soc.", 2009, vol. 41, n<sup>o</sup> 2, pp. 253–273, <http://dx.doi.org/10.1112/blms/bdn122>
- [49] T. BOURGERON, Z. XU, M. DOUMIC, M. T. TEIXEIRA. *The asymmetry of telomere replication contributes to replicative senescence heterogeneity*, in "Scientific reports", 2015, vol. 5
- [50] A. CHRONOPOULOU, S. TINDEL. *On inference for fractional differential equations*, in "Statistical Inference for Stochastic Processes", 2013, vol. 16, n<sup>o</sup> 1, pp. 29–61
- [51] J. F. COEURJOLLY. *Simulation and identification of the fractional brownian motion: a bibliographical and comparative study*, in "Journal of Statistical Software", 2000, vol. 5, pp. 1–53
- [52] S. COHEN, C. LACAUX, M. LEDOUX. *A general framework for simulation of fractional fields*, in "Stochastic Process. Appl.", 2008, vol. 118, n<sup>o</sup> 9, pp. 1489–1517, <http://dx.doi.org/10.1016/j.spa.2007.09.008>
- [53] M. H. DAVIS. *Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models*, in "Journal of the Royal Statistical Society. Series B (Methodological)", 1984, pp. 353–388

- [54] A. DEYA, S. TINDEL. *Rough Volterra equations. I. The algebraic integration setting*, in "Stoch. Dyn.", 2009, vol. 9, n<sup>o</sup> 3, pp. 437–477, <http://dx.doi.org/10.1142/S0219493709002737>
- [55] J. A. DOORNIK, H. HANSEN. *An omnibus test for univariate and multivariate normality\**, in "Oxford Bulletin of Economics and Statistics", 2008, vol. 70, n<sup>o</sup> s1, pp. 927–939
- [56] M. DOUMIC, M. HOFFMANN, N. KRELL, L. ROBERT. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*, in "Bernoulli", 2015, vol. 21, n<sup>o</sup> 3, pp. 1760–1799
- [57] J. FEARNLEY. *Exponential lower bounds for policy iteration*, in "37th international colloquium conference on Automata, languages and programming: Part II", Berlin, Heidelberg, ICALP'10, Springer-Verlag, 2010, pp. 551–562
- [58] S. FERRIGNO, G. R. DUCHARME. *Un test d'adéquation global pour la fonction de répartition conditionnelle*, in "Comptes Rendus Mathématique", 2005, vol. 341, n<sup>o</sup> 5, pp. 313–316
- [59] S. FERRIGNO, M. MAUMY-BERTRAND, A. MULLER. *Uniform law of the logarithm for the local linear estimator of the conditional distribution function*, in "C. R. Math. Acad. Sci. Paris", 2010, vol. 348, n<sup>o</sup> 17-18, pp. 1015–1019, <http://dx.doi.org/10.1016/j.crma.2010.08.003>
- [60] J. FRIEDMAN, T. HASTIE, R. TIBSHIRANI. *Sparse inverse covariance estimation with the graphical lasso*, in "Biostatistics", 2008, vol. 9, n<sup>o</sup> 3, pp. 432–441
- [61] A. GENADOT, M. THIEULLEN. *Averaging for a fully coupled piecewise-deterministic markov process in infinite dimensions*, in "Advances in Applied Probability", 2012, vol. 44, n<sup>o</sup> 3, pp. 749–773
- [62] C. GIRAUD, S. HUET, N. VERZELEN. *Graph selection with GGMselect*, in "Statistical applications in genetics and molecular biology", 2012, vol. 11, n<sup>o</sup> 3
- [63] M. GRABISCH, T. MUROFUSHI, M. SUGENO. *Fuzzy measure of fuzzy events defined by fuzzy integrals*, in "Fuzzy Sets and Systems", 1992, vol. 50, n<sup>o</sup> 3, pp. 293–313
- [64] M. GUBINELLI, S. TINDEL. *Rough evolution equations*, in "Ann. Probab.", 2010, vol. 38, n<sup>o</sup> 1, pp. 1–75, <http://dx.doi.org/10.1214/08-AOP437>
- [65] S. J. HABERMAN. *Algorithm AS 51: Log-linear fit for contingency tables*, in "Applied Statistics", 1972, pp. 218–225
- [66] T. HANSEN, U. ZWICK. *Lower Bounds for Howard's Algorithm for Finding Minimum Mean-Cost Cycles*, in "ISAAC (1)", 2010, pp. 415-426
- [67] S. HERRMANN, P. VALLOIS. *From persistent random walk to the telegraph noise*, in "Stoch. Dyn.", 2010, vol. 10, n<sup>o</sup> 2, pp. 161–196, <http://dx.doi.org/10.1142/S0219493710002905>
- [68] J. HU, W.-C. WU, S. SASTRY. *Modeling subtilin production in bacillus subtilis using stochastic hybrid systems*, in "Hybrid Systems: Computation and Control", Springer, 2004, pp. 417–431

- [69] R. KEINJ, T. BASTOGNE, P. VALLOIS. *Multinomial model-based formulations of TCP and NTCP for radiotherapy treatment planning*, in "Journal of Theoretical Biology", June 2011, vol. 279, n<sup>o</sup> 1, pp. 55-62 [DOI : 10.1016/J.JTBI.2011.03.025], <http://hal.inria.fr/hal-00588935/en>
- [70] R. KOENKER. *Quantile regression*, Cambridge university press, 2005, n<sup>o</sup> 38
- [71] H. J. KUSHNER, G. YIN. *Stochastic approximation and recursive algorithms and applications*, Springer Science & Business Media, 2003, vol. 35
- [72] Y. A. KUTOYANTS. *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London Ltd., London, 2004, xiv+481 p.
- [73] C. LACAUX. *Real Harmonizable Multifractional Lévy Motions*, in "Ann. Inst. Poincaré.", 2004, vol. 40, n<sup>o</sup> 3, pp. 259–277
- [74] C. LACAUX. *Series representation and simulation of multifractional Lévy motions*, in "Adv. in Appl. Probab.", 2004, vol. 36, n<sup>o</sup> 1, pp. 171–197, <http://dx.doi.org/10.1239/aap/1077134469>
- [75] C. LACAUX, J.-M. LOUBES. *Hurst exponent estimation of fractional Lévy motion*, in "ALEA Lat. Am. J. Probab. Math. Stat.", 2007, vol. 3, pp. 143–164
- [76] C. LACAUX, R. MARTY. *From invariance principles to a class of multifractional fields related to fractional sheets*, 2011, MAP5 2011-08, <http://hal.archives-ouvertes.fr/hal-00592188/en/>
- [77] M. LAVIELLE. *Mixed effects models for the population approach: models, tasks, methods and tools*, CRC Press, 2014
- [78] L. LEBART. *On the Benzecri's method for computing eigenvectors by stochastic approximation (the case of binary data)*, in "Compstat 1974 (Proc. Sympos. Computational Statist., Univ. Vienna, Vienna, 1974)", Vienna, Physica Verlag, 1974, pp. 202–211
- [79] B. LESNER, B. SCHERRER. *Non-Stationary Approximate Modified Policy Iteration*, in "ICML 2015", Lille, France, July 2015, <https://hal.inria.fr/hal-01186664>
- [80] T. LYONS, Z. QIAN. *System control and rough paths*, Oxford mathematical monographs, Clarendon Press, 2002, <http://books.google.com/books?id=H9fRQNIngZYC>
- [81] J.-L. MARICHAL. *An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria*, in "Fuzzy Systems, IEEE Transactions on", 2000, vol. 8, n<sup>o</sup> 6, pp. 800–807
- [82] N. MEINSHAUSEN, P. BÜHLMANN. *High-dimensional graphs and variable selection with the lasso*, in "The Annals of Statistics", 2006, pp. 1436–1462
- [83] J.-M. MONNEZ. *Approximation stochastique en analyse factorielle multiple*, in "Ann. I.S.U.P.", 2006, vol. 50, n<sup>o</sup> 3, pp. 27–45
- [84] J.-M. MONNEZ. *Stochastic approximation of the factors of a generalized canonical correlation analysis*, in "Statist. Probab. Lett.", 2008, vol. 78, n<sup>o</sup> 14, pp. 2210–2216, <http://dx.doi.org/10.1016/j.spl.2008.01.088>

- [85] J.-M. MONNEZ. *Convergence d'un processus d'approximation stochastique en analyse factorielle*, in "Publ. Inst. Statist. Univ. Paris", 1994, vol. 38, n<sup>o</sup> 1, pp. 37–55
- [86] E. NADARAYA. *On non-parametric estimates of density functions and regression curves*, in "Theory of Probability & Its Applications", 1965, vol. 10, n<sup>o</sup> 1, pp. 186–190
- [87] A. NEUENKIRCH, I. NOURDIN, A. RÖSSLER, S. TINDEL. *Trees and asymptotic expansions for fractional stochastic differential equations*, in "Ann. Inst. Henri Poincaré Probab. Stat.", 2009, vol. 45, n<sup>o</sup> 1, pp. 157–174, <http://dx.doi.org/10.1214/07-AIHP159>
- [88] A. NEUENKIRCH, S. TINDEL. *A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise*, in "Statistical Inference for Stochastic Processes", 2014, vol. 17, n<sup>o</sup> 1, pp. 99–120
- [89] I. POST, Y. YE. *The simplex method is strongly polynomial for deterministic Markov decision processes*, arXiv:1208.5083v2, 2012
- [90] M. PUTERMAN. *Markov Decision Processes*, Wiley, New York, 1994
- [91] Q. QI, J. A. WATTIS, H. M. BYRNE. *Stochastic simulations of normal aging and Werner's syndrome*, in "Bulletin of mathematical biology", 2014, vol. 76, n<sup>o</sup> 6, pp. 1241–1269
- [92] P. RETIF. *Modeling, digital simulation and analysis of nanoparticles-X ray interaction. Applications to augmented radiotherapy.*, Université de Lorraine, March 2016, <https://hal.archives-ouvertes.fr/tel-01376033>
- [93] B. ROYNETTE, P. VALLOIS, M. YOR. *Brownian penalizations related to excursion lengths, VII*, in "Annales de l'IHP Probabilités et statistiques", 2009, vol. 45, n<sup>o</sup> 2, pp. 421–452
- [94] F. RUSSO, P. VALLOIS. *Stochastic calculus with respect to continuous finite quadratic variation processes*, in "Stochastics: An International Journal of Probability and Stochastic Processes", 2000, vol. 70, n<sup>o</sup> 1-2, pp. 1–40
- [95] F. RUSSO, P. VALLOIS. *Elements of stochastic calculus via regularization*, in "Séminaire de Probabilités XL", Berlin, Lecture Notes in Math., Springer, 2007, vol. 1899, pp. 147–185, [http://dx.doi.org/10.1007/978-3-540-71189-6\\_7](http://dx.doi.org/10.1007/978-3-540-71189-6_7)
- [96] P. SALMINEN, P. VALLOIS. *On maximum increase and decrease of Brownian motion*, in "Annales de l'Institut Henri Poincaré (B) Probability and Statistics", Elsevier, 2007, vol. 43, n<sup>o</sup> 6, pp. 655–676
- [97] B. SCHERRER, M. GHAVAMZADEH, V. GABILLON, B. LESNER, M. GEIST. *Approximate Modified Policy Iteration and its Application to the Game of Tetris*, in "Journal of Machine Learning Research", 2015, vol. 16, pp. 1629–1676, A paraître, <https://hal.inria.fr/hal-01091341>
- [98] B. SCHERRER, B. LESNER. *On the Use of Non-Stationary Policies for Stationary Infinite-Horizon Markov Decision Processes*, in "NIPS 2012 - Neural Information Processing Systems", South Lake Tahoe, United States, December 2012, <https://hal.inria.fr/hal-00758809>

- 
- [99] B. SCHERRER. *Performance Bounds for Lambda Policy Iteration and Application to the Game of Tetris*, in "Journal of Machine Learning Research", January 2013, vol. 14, pp. 1175-1221, <https://hal.inria.fr/hal-00759102>
- [100] B. SCHERRER. *Approximate Policy Iteration Schemes: A Comparison*, in "ICML - 31st International Conference on Machine Learning - 2014", Pékin, China, June 2014, <https://hal.inria.fr/hal-00989982>
- [101] B. SCHERRER. *Improved and Generalized Upper Bounds on the Complexity of Policy Iteration*, in "Mathematics of Operations Research", 2015, A paraître
- [102] L. SHAPLEY. *Stochastic Games*, in "Proceedings of the National Academy of Sciences", 1953, vol. 39, n<sup>o</sup> 10, pp. 1095-1100
- [103] C. A. TUDOR, F. G. VIENS. *Statistical aspects of the fractional stochastic calculus*, in "Ann. Statist.", 2007, vol. 35, n<sup>o</sup> 3, pp. 1183–1212, <http://dx.doi.org/10.1214/009053606000001541>
- [104] P. VALLOIS, C. S. TAPIERO. *Memory-based persistence in a counting random walk process*, in "Phys. A.", 2007, vol. 386, n<sup>o</sup> 1, pp. 303–307, <http://dx.doi.org/10.1016/j.physa.2007.08.027>
- [105] P. VALLOIS. *The range of a simple random walk on  $\mathbb{Z}$* , in "Advances in applied probability", 1996, pp. 1014–1033
- [106] N. VILLA-VIALANEIX. *An introduction to network inference and mining*, 2015, <http://wikistat.fr/>, (consulté le 22/07/2015), [http://www.nathalievilla.org/doc/pdf/wikistat-network\\_compiled.pdf](http://www.nathalievilla.org/doc/pdf/wikistat-network_compiled.pdf)
- [107] Z. XU, K. D. DUC, D. HOLCMAN, M. T. TEIXEIRA. *The length of the shortest telomere as the major determinant of the onset of replicative senescence*, in "Genetics", 2013, vol. 194, n<sup>o</sup> 4, pp. 847–857
- [108] Y. YE. *The Simplex and Policy-Iteration Methods Are Strongly Polynomial for the Markov Decision Problem with a Fixed Discount Rate*, in "Math. Oper. Res.", 2011, vol. 36, n<sup>o</sup> 4, pp. 593-603