

*Inria*

IN PARTNERSHIP WITH:  
**CNRS**

**Université de Lorraine**

Activity Report 2019

# 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. Team, Visitors, External Collaborators</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>2</b>
<b>3. Research Program</b>	<b>2</b>
3.1. Introduction	2
3.2. Stochastic modeling	3
3.3. Estimation and control for stochastic processes	3
3.4. Algorithms and estimation for graph data	3
3.5. Regression and machine learning	4
<b>4. Application Domains</b>	<b>4</b>
4.1. Tumor growth-oncology	4
4.2. Genomic data and micro-organisms population	4
4.3. Epidemiology and e-health	5
4.4. Dynamics of telomeres	5
<b>5. Highlights of the Year</b>	<b>5</b>
<b>6. New Software and Platforms</b>	<b>5</b>
6.1. Angio-Analytics	5
6.2. ARMADA	6
6.3. kosel	6
6.4. SesIndexCreatoR	7
6.5. In silico	7
6.6. HSPOR	7
<b>7. New Results</b>	<b>7</b>
7.1. Stochastic modelling	7
7.2. Optimal Control of Markov Processes	8
7.3. Algorithms and Estimation for graph data	8
7.4. Regression and machine learning	9
<b>8. Bilateral Contracts and Grants with Industry</b>	<b>11</b>
<b>9. Partnerships and Cooperations</b>	<b>11</b>
9.1. Regional Initiatives	11
9.2. National Initiatives	11
9.3. International Research Visitors	12
<b>10. Dissemination</b>	<b>12</b>
10.1. Promoting Scientific Activities	12
10.1.1. Scientific Events: Organisation	12
10.1.2. Scientific Events: Selection	12
10.1.2.1. Chair of Conference Program Committees	12
10.1.2.2. Member of the Conference Program Committees	12
10.1.3. Journal	12
10.1.4. Invited Talks	12
10.1.5. Research Administration	12
10.2. Teaching - Supervision - Juries	13
10.2.1. Teaching	13
10.2.2. Supervision	14
10.2.3. Juries	14
10.3. Popularization	14
<b>11. Bibliography</b>	<b>15</b>



## Project-Team BIGS

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

### Keywords:

#### Computer Science and Digital Science:

- A3.1. - Data
  - A3.1.1. - Modeling, representation
- A3.2. - Knowledge
  - A3.2.3. - Inference
- A3.3. - Data and knowledge analysis
  - A3.3.1. - On-line analytical processing
  - A3.3.2. - Data mining
  - A3.3.3. - Big data analysis
- A3.4.1. - Supervised learning
- A3.4.2. - Unsupervised learning
- A3.4.4. - Optimization and learning
- A3.4.7. - Kernel methods
- A6. - Modeling, simulation and control
  - A6.1. - Methods in mathematical modeling
    - A6.1.2. - Stochastic Modeling
  - A6.2. - Scientific computing, Numerical Analysis & Optimization
    - A6.2.3. - Probabilistic methods
    - A6.2.4. - Statistical methods
  - A6.4. - Automatic control
    - A6.4.2. - Stochastic control

#### Other Research Topics and Application Domains:

- B1. - Life sciences
  - B1.1. - Biology
    - B1.1.2. - Molecular and cellular biology
    - B1.1.10. - Systems and synthetic biology
    - B1.1.11. - Plant Biology
  - B2.2. - Physiology and diseases
    - B2.2.1. - Cardiovascular and respiratory diseases
    - B2.2.3. - Cancer
  - B2.3. - Epidemiology
  - B2.4. - Therapies
- B5.5. - Materials

## 1. Team, Visitors, External Collaborators

### Research Scientist

Bruno Scherrer [Inria, Researcher, HDR]

**Faculty Members**

Anne Gégout Petit [Team leader, Univ de Lorraine, Professor, HDR]  
Thierry Bastogne [Univ de Lorraine, Associate Professor, HDR]  
Sandie Ferrigno [Univ de Lorraine, Associate Professor]  
Sophie Mezieres [Univ de Lorraine, Associate Professor]  
Jean-Marie Monnez [Univ de Lorraine, Emeritus Professor, HDR]  
Aurélien Muller-Gueudin [Univ de Lorraine, Associate Professor]  
Pierre Vallois [Univ de Lorraine, Professor, HDR]

**Post-Doctoral Fellows**

Lionel Lenôtre [Univ de Lorraine, until Aug 2019]  
Emma Horton [from Dec 2019]

**PhD Students**

Florine Greciet [SAFRAN Aircraft Engines]  
Pauline Guyot [Univ de Lorraine, until Nov 2019]  
Clémence Karmann [Inria]  
Nassim Sahki [Inria]

**Technical staff**

Benoît Lalloué [Univ de Lorraine, Engineer]

**Interns and Apprentices**

Olivier Coudray [Univ de Lorraine, from Apr 2019 until Aug 2019]  
Christophe Reype [Univ de Strasbourg, from Mar 2019 until Sep 2019]

**Administrative Assistant**

Céline Cordier [Inria]

## 2. Overall Objectives

### 2.1. Overall Objectives

BIGS is a joint team of Inria, CNRS and Université Lorraine, via the Institut Élie Cartan, UMR 7502 CNRS-UL laboratory in mathematics, of which Inria is a strong partner. One member of BIGS, T. Bastogne, comes from the Research Center of Automatic Control of Nancy (CRAN), with which BIGS has strong relations in the domain "Health-Biology-Signal". Our research is mainly focused on stochastic modeling and statistics but also aiming at a better understanding of biological systems. BIGS involves 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 issues arising in life sciences, with a special focus on (1) tumor growth, (2) photodynamic therapy, (3) population studies of genomic data and of micro-organisms genomics, (4) epidemiology and e-health.

## 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 clarity, 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 [80], [57]. Piecewise deterministic processes are non diffusion processes also frequently used in the biological context [47], [56], [49]. Among Markov model, we developed strong expertise about processes derived from Brownian motion and Stochastic Differential Equations [72], [55]. For instance, knowledge about Brownian or random walk excursions [79], [71] 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 fractional fields. The basic aspects of these differential equations are now well understood, mainly thanks to the so-called rough paths tools [63], but also invoking the Russo-Vallois integration techniques [73]. The specific issue of Volterra equations driven by fractional Brownian motion, which is central for the subdiffusion within proteins problem, is addressed in [48]. Many generalizations (Gaussian or not) of this model have been recently proposed for some Gaussian locally self-similar fields, or for some non-Gaussian models [60], or for anisotropic models [44].

## 3.3. Estimation and control for stochastic processes

We develop inference about stochastic processes that we use for modeling. Control of stochastic processes is also a way to optimise administration (dose, frequency) of therapy.

There are many estimation techniques for diffusion processes or coefficients of fractional or multifractional Brownian motion according to a set of observations [59], [40], [46]. But, the inference problem for diffusions driven by a fractional Brownian motion is still in its infancy. Our team has a good expertise about inference of the jump rate and the kernel of Piecewise Deterministic Markov Processes (PDMP) [37], [38], [36], [39]. 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 tackle the problem of inference of "Hidden PDMP". As an example, in pharmacokinetics modeling inference, we want to take into account for presence of timing noise and identification from longitudinal data. We have expertise on this subjects [41], and we also used mixed models to estimate tumor growth [42].

We consider the control of stochastic processes within the framework of Markov Decision Processes [70] and their generalization known as multi-player stochastic games, 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 [43]. Regarding complexity, a central topic of research is the analysis of the Policy Iteration algorithm, which has made significant progress in the last years [82], [69], [54], [78], 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 [76], [75], [77], [62], [74], 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 pairs of these nodes called edges. This type of data is frequently used in biology because they provide a mathematical representation of many concepts such as biological structures and networks of relationships in a population. Some attention has recently been focused in the group on modeling and inference for graph data.

Network inference is the process of making inference about the link between two variables taking into account the information about other variables. [81] gives a very good introduction and many references about network inference and mining. Many methods are available to infer and test edges in Gaussian graphical models [81], [64], [52], [53]. However, when dealing with abundance data, because inflated zero data, we are far from gaussian assumption and we want to develop inference in this case.

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. We also work on lossy compression of trees via directed acyclic graphs. These methods enable us to compute distances between tree data faster than from the original structures and with a high accuracy.

### 3.5. Regression and machine learning

Regression models and machine learning aim at inferring statistical links between a variable of interest and covariates. In biological study, it is always important to develop adapted learning methods both in the context of *standard* data and also for data of high dimension (with sometimes few observations) and very massive or online data.

Many methods are available to estimate conditional quantiles and test dependencies [68], [58]. Among them we have developed nonparametric estimation by local analysis via kernel methods [50], [51] and we want to study properties of this estimator in order to derive a 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 develop omnibus tests that examine several assumptions together.

Concerning the analysis of high dimensional data, our view on the topic relies on the *French data analysis school*, specifically on Factorial Analysis tools. In this context, stochastic approximation is an essential tool [61], which allows one to approximate eigenvectors in a stepwise manner [67], [65], [66]. BIGS aims at performing accurate classification or clustering by taking advantage of the possibility of updating the information "online" using stochastic approximation algorithms [45]. We focus on several incremental procedures for regression and data analysis like linear and logistic regressions and PCA (Principal Component Analysis).

We also focus on the biological context of high-throughput bioassays in which several hundreds or thousands of biological signals are measured for a posterior analysis. We have to account for the inter-individual variability within the modeling procedure. 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-oncology

On this topic, we want to propose branching processes to model appearance of mutations in tumor through new collaborations with clinicians. The observed process is the "circulating DNA" (ctDNA). The final purpose is to use ctDNA as a early biomarker of the resistance to an immunotherapy treatment. It is the aim of the ITMO project. Another topic is the identification of dynamic network of expression. In the ongoing work on low-grade gliomas, a local database of 400 patients will be soon available to construct models. We plan to extend it through national and international collaborations (Montpellier CHU, Montreal CRHUM). Our aim is to build a decision-aid tool for personalised medicine. In the same context, there is a topic of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

### 4.2. Genomic data and micro-organisms population

Despite of his 'G' in the name of BIGS, Genetics is not central in the applications of the team. However, we want to contribute to a better understanding of the correlations between genes trough their expression data and of the genetic bases of drug response and disease. We have contributed to methods detecting proteomics and transcriptomics variables linked with the outcome of a treatment.



### 4.3. Epidemiology and e-health

We have many works to do in our ongoing projects in the context of personalized medicine with CHU Nancy. They deal with biomarkers research, prognostic value of quantitative variables and events, scoring, and adverse events. We also want to develop our expertise in rupture detection in a project with APHP (Assistance Publique Hôpitaux de Paris) for the detection of adverse events, earlier than the clinical signs and symptoms. The clinical relevance of predictive analytics is obvious for high-risk patients such as those with solid organ transplantation or severe chronic respiratory disease for instance. The main challenge is the rupture detection in multivariate and heterogeneous signals (for instance daily measures of electrocardiogram, body temperature, spirometry parameters, sleep duration, etc). Other collaborations with clinicians concern foetopathology and we want to use our work on conditional distribution function to explain fetal and child growth. We have data from the "Service de foetopathologie et de placentologie" of the "Maternité Régionale Universitaire" (CHU Nancy).

### 4.4. Dynamics of telomeres

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. Through a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently obtained data on the distribution of the length of telomeres from blood cells. With members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

## 5. Highlights of the Year

### 5.1. Highlights of the Year

#### 5.1.1. Awards

BIGS participated to the organization the JdS 2019 (Journées de Statistique 2019) <http://www.jds2019.sfds.asso.fr/> in Nancy.

B. Scherrer and his co-authors received the Outstanding paper award for AAAI-2019“ from the AAAI (Association for the Advancement of Artificial Intelligence).

BEST PAPER AWARD:

[9]

Y. EFRONI, G. DALAL, B. SCHERRER, S. MANNOR. *How to Combine Tree-Search Methods in Reinforcement Learning*, in "AAAI 19 - Thirty-Third AAAI Conference on Artificial Intelligence", Honolulu, Hawaii, United States, January 2019, <https://arxiv.org/abs/1809.01843> - AAAI 2019, <https://hal.inria.fr/hal-02273713>

## 6. New Software and Platforms

### 6.1. Angio-Analytics

KEYWORDS: Health - Cancer - Biomedical imaging

**SCIENTIFIC DESCRIPTION:** This tool allows the pharmacodynamic characterization of anti-vascular effects in anti-cancer treatments. It uses time series of in vivo images provided by intra-vital microscopy. Such in vivo images are obtained owing to skinfold chambers placed on mice skin. The automatized analysis is split up into two steps that were completely performed separately and manually before. The first steps corresponds to image processing to identify characteristics of the vascular network. The last step is the system identification of the pharmacodynamic response and the statistical analysis of the model parameters.

**FUNCTIONAL DESCRIPTION:** Angio-Analytics allows the pharmacodynamic characterization of anti-vascular effects in anti-cancer treatments.

- Participant: Thierry Bastogne
- Contact: Thierry Bastogne

## 6.2. ARMADA

*A Statistical Methodology to Select Covariates in High-Dimensional Data under Dependence*

**KEYWORDS:** Biostatistics - Aggregated methods - High Dimensional Data - Personalized medicine - Variable selection

**FUNCTIONAL DESCRIPTION:** Two steps variable selection procedure in a context of high-dimensional dependent data but few observations. First step is dedicated to eliminate dependence between variables (clustering of variables, followed by factor analysis inside each cluster). Second step is a variable selection using by aggregation of adapted methods. <https://hal.archives-ouvertes.fr/hal-02173568>

**NEWS OF THE YEAR:** This package is a new one.

- Participants: Aurélie Muller and Anne Gégout-Petit
- Contact: Aurélie Muller
- Publication: [Package 'armada' : A Statistical Methodology to Select Covariates in High-Dimensional Data under Dependence](#)
- URL: <https://cran.r-project.org/web/packages/armada/>

## 6.3. kosel

*Variable Selection by Revisited Knockoffs Procedures*

**KEYWORDS:** Variable selection - Regression

**FUNCTIONAL DESCRIPTION:** Performs variable selection for many types of L1-regularised regressions using the revisited knockoffs procedure. This procedure uses a matrix of knockoffs of the covariates independent from the response variable Y. The idea is to determine if a covariate belongs to the model depending on whether it enters the model before or after its knockoff. The procedure suits for a wide range of regressions with various types of response variables. Regression models available are exported from the R packages 'glmnet' and 'ordinalNet'. Based on the paper linked to via the URL below: Gegout A., Gueudin A., Karmann C. (2019) <https://arxiv.org/abs/1907.03153>

**NEWS OF THE YEAR:** This package is a new one.

- Participants: Clémence Karmann, Aurélie Muller and Anne Gégout-Petit
- Contact: Aurélie Muller
- Publication: [The revisited knockoffs method for variable selection in  \$L\_1\$ -penalised regressions.](#)
- URL: <https://cran.r-project.org/web/packages/kosel/kosel.pdf>

## 6.4. SesIndexCreator

**FUNCTIONAL DESCRIPTION:** This package allows computing and visualizing socioeconomic indices and categories distributions from datasets of socioeconomic variables (These tools were developed as part of the EquitArea Project, a public health program).

- Participants: Benoît Lalloué, Jean-Marie Monnez, Nolwenn Le Meur and Severine Deguen
- Contact: Benoît Lalloué
- URL: [http://www.equitarea.org/documents/packages\\_1.0-0/](http://www.equitarea.org/documents/packages_1.0-0/)

## 6.5. In silico

*In silico design of nanoparticles for the treatment of cancers by enhanced radiotherapy*

**KEYWORDS:** Bioinformatics - Cancer - Drug development

**FUNCTIONAL DESCRIPTION:** To speed up the preclinical development of medical engineered nanomaterials, we have designed an integrated computing platform dedicated to the virtual screening of nanostructured materials activated by X-ray making it possible to select nano-objects presenting interesting medical properties faster. The main advantage of this in silico design approach is to virtually screen a lot of possible formulations and to rapidly select the most promising ones. The platform can currently handle the accelerated design of radiation therapy enhancing nanoparticles and medical imaging nano-sized contrast agents as well as the comparison between nano-objects and the optimization of existing materials.

- Participant: Thierry Bastogne
- Contact: Thierry Bastogne

## 6.6. HSPOR

*Hidden Smooth Polynomial Regression for Rupture Detection*

**KEYWORDS:** Polynomial regression - Rupture detection

**FUNCTIONAL DESCRIPTION:** Several functions that allow by different methods to infer a piecewise polynomial regression model under regularity constraints, namely continuity or differentiability of the link function. The implemented functions are either specific to data with two regimes, or generic for any number of regimes, which can be given by the user or learned by the algorithm.

**NEWS OF THE YEAR:** This package is a new one

- Participants: Florine Greciet, Romain Azais and Anne Gégout-Petit
- Contact: Florine Greciet
- URL: <https://cran.r-project.org/web/packages/HSPOR/>

# 7. New Results

## 7.1. Stochastic modelling

Participants: A. Gégout-Petit, S. Mézières, P. Vallois

In the framework of the esca-illness of vines, we 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 (2), we developed an autologistic model (centered in a new way), have proposed estimators of the parameters and showed their properties and proposed a way to choose between several neighborhood models. These results were published in Spatial Statistics [4].

In a collaboration with physicists from Nancy CHRU, we have worked about the interest to use the whole distribution of telomeres lengths until the mean that is usually used to characterise ageing of a cell. We have shown that the shape of the distribution can be seen as a individuals's signature. It is the object of the paper published in Scientific Reports [8].

After preliminary suggestions on the building of models for low-grade gliomas [3], we focused our attention on the diffuse character of such tumors. We characterized the infiltrating phenotype (infiltration rate, direction of infiltration, evolution of morphology over time) as a new variable to consider in a context of multifactorial modelling (submitted article). A monocentric retrospective study has been conducted on the local database, estimating survival paramaters and comparing the effects of treatments (writing article). A brain cartography obtained by sensorial simulations during awake surgery with the aid of clustering analysis has been published in "Brain - A Journal of Neurology" [6].

## 7.2. Optimal Control of Markov Processes

Participants: B. Scherrer

Finite-horizon lookahead policies are abundantly used in Reinforcement Learning and demonstrate impressive empirical success. Usually, the lookahead policies are implemented with specific planning methods such as Monte Carlo Tree Search (e.g. in AlphaZero). Referring to the planning problem as tree search, a reasonable practice in these implementations is to back up the value only at the leaves while the information obtained at the root is not leveraged other than for updating the policy. Here, we question the potency of this approach. Namely, the latter procedure is non-contractive in general, and its convergence is not guaranteed. Our proposed enhancement, in [9], published in AAAI'2019, is straightforward and simple: use the return from the optimal tree path to back up the values at the descendants of the root. This leads to a  $\gamma^h$ -contracting procedure, where  $\gamma$  is the discount factor and  $h$  is the tree depth. To establish our results, we first introduce a notion called *multiple-step greedy consistency*. We then provide convergence rates for two algorithmic instantiations of the above enhancement in the presence of noise injected to both the tree search stage and value estimation stage.

Value iteration is a method to generate optimal control inputs for generic nonlinear systems and cost functions. Its implementation typically leads to approximation errors, which may have a major impact on the closed-loop system performance. We talk in this case of approximate value iteration (AVI). In [24], published in CDC'2019, we investigate the stability of systems for which the inputs are obtained by AVI. We consider deterministic discrete-time nonlinear plants and a class of general, possibly discounted, costs. We model the closed-loop system as a family of systems parameterized by tunable parameters, which are used for the approximation of the value function at different iterations, the discount factor and the iteration step at which we stop running the algorithm. It is shown, under natural stabilizability and detectability properties as well as mild conditions on the approximation errors, that the family of closed-loop systems exhibit local practical stability properties. The analysis is based on the construction of a Lyapunov function given by the sum of the approximate value function and the Lyapunov-like function that characterizes the detectability of the system. By strengthening our conditions, asymptotic and exponential stability properties are guaranteed.

Many recent successful (deep) reinforcement learning algorithms make use of regularization, generally based on entropy or Kullback-Leibler divergence. In [10], published in ICML'2019, we propose a general theory of regularized Markov Decision Processes that generalizes these approaches in two directions: we consider a larger class of regularizers, and we consider the general modified policy iteration approach, encompassing both policy iteration and value iteration. The core building blocks of this theory are a notion of regularized Bellman operator and the Legendre-Fenchel transform, a classical tool of convex optimization. This approach allows for error propagation analyses of general algorithmic schemes of which (possibly variants of) classical algorithms such as Trust Region Policy Optimization, Soft Q-learning, Stochastic Actor Critic or Dynamic Policy Programming are special cases. This also draws connections to proximal convex optimization, especially to Mirror Descent.

## 7.3. Algorithms and Estimation for graph data

Participants: A. Gégout-Petit, A. Gueudin, C. Karmann

We consider the problem of graph estimation in a zero-inflated Gaussian model. In this model, zero-inflation is obtained by double truncation (right and left) of a Gaussian vector. The goal is to recover the latent graph structure of the Gaussian vector with observations of the zero-inflated truncated vector. We propose a two step estimation procedure. The first step consists in estimating each term of the covariance matrix by maximising the corresponding bivariate marginal log-likelihood of the truncated vector. The second one uses the graphical lasso procedure to estimate the sparsity of the precision matrix, which encodes the graph structure. We then state some theoretical results about the convergence rate of the covariance matrix and precision matrix estimators. These results allow us to establish consistency of our procedure with respect to graph structure recovery. We also present some simulation studies to corroborate the efficiency of our procedure. It is the object of the submitted paper [29], a part of the PhD thesis [1] and the communications [16] [15].

## 7.4. Regression and machine learning

Participants: E. Albuissou, T. Bastogne, S. Ferrigno, A. Gégout-Petit, F. Greciet, P. Guyot, C. Karmann, J.-M. Monnez, N. Sahki, S. Mézières, B. Lalloué

Through a collaboration with the pharmaceutical company Transgene (Strasbourg, France), we have developed a method for selecting covariates. The problem posed by Transgene was to establish patient profiles on the basis of their response to a treatment developed by Transgene. We have then proposed a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. A simulation study shows the interest of the decorrelation inside the different clusters of covariates. We have applied our method to the data of Transgene. For instance, transcriptomic data of 37 patients with advanced non-small-cell lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment. Our method has also been applied in another collaboration with biologists (CRAN laboratory, Nancy, France). In that case, our method has been applied to transcriptomic data of 79 breast tumor samples, to define patient profiles for a new metastatic biomarker and associated gene network. Our developed method is a contribution to the development of personalized medicine. We have published the method, as well as the two applications in [27].

In order to detect change of health state for lung-transplanted patient, we have begun to work on breakdowns in multivariate physiological signals. We consider the score-based CUSUM statistic and propose to evaluate the detection performance of some thresholds on simulation data. Two thresholds come from the literature: Wald's constant and Margavio's instantaneous threshold, and three contribution thresholds built by a simulation-based procedure: the first one is constant, the second instantaneous and the third is a dynamical version of the previous one. The threshold's performance is evaluated for several scenarii, according to the detection objective and the real change in the data. The simulation results show that Margavio's threshold is the best at minimizing the detection delay while maintaining the given false alarm rate. But on real data, we suggest to use the dynamic instantaneous threshold because it is the easiest to build for practical implementation. It is the purpose of the communication [11] and the submitted paper [35].

We consider the problem of variable selection in regression models. In particular, we are interested in selecting explanatory covariates linked with the response variable and we want to determine which covariates are relevant, that is which covariates are involved in the model. In this framework, we deal with L1-penalised regression models. To handle the choice of the penalty parameter to perform variable selection, we develop a new method based on knockoffs. This revisited knockoffs method is general, suitable for a wide range of regressions with various types of response variables. Besides, it also works when the number of observations is smaller than the number of covariates and gives an order of importance of the covariates. Finally, we provide many experimental results to corroborate our method and compare it with other variable selection methods. It is the object of communication [17], the submitted paper [30] and a chapter of the PhD thesis [1].

In order to model crack propagation rate, continuous physical phenomenon that presents several regimes, we proposed a piecewise polynomial regression model under continuity and/or derivability assumptions as well as a statistical inference method to estimate the transition times and the parameters of each regime. We proposed

several algorithms and studied their efficiency. The most efficient algorithm relies on dynamic programming. It is the object of the communication [14] and the PhD thesis of Florine Greciet.

Let consider a regression model  $Y = m(X) + \sigma(X)\varepsilon$  to explain  $Y$  from  $X$ , where  $m(\cdot)$  is the regression function,  $\sigma^2(\cdot)$  the variance function and  $\varepsilon$  the random error term. Methods to assess how well a model fits a set of observations fall under the banner of goodness-of-fit tests. Many tests have been developed to assess the different assumptions for this kind of model. Most of them are “directional” in that they detect departures from mainly a given assumption of the model. Other tests are “global” in that they assess whether a model fits a data set on all its assumptions. We focus on the task of choosing the structural part  $m(\cdot)$ . It gets most attention because it contains easily interpretable information about the relationship between  $X$  and  $Y$ . To validate the form of the regression function, we consider three nonparametric tests based on a generalization of the Cramér-von Mises statistic. The first two are directional tests, while the third is a global test. To perform these goodness-of-fit tests based on a generalization of the Cramér-von Mises statistic, we have used Wild bootstrap methods and we also proposed a method to choose the bandwidth parameter used in nonparametric estimations. Then, we have developed the `cvmgof` R package, an easy-to-use tool for many users. The use of the package is described and illustrated using simulations to compare the three implemented tests in a paper in progress.

In epidemiology, we are working with INSERM clinicians and biostatisticians to study fetal development in the last two trimesters of pregnancy. Reference or standard curves are required in this kind of biomedical problems. Values which lie outside the limits of these reference curves may indicated the presence of disorder. Data are from the French EDEN mother-child cohort (INSERM). It is a mother-child cohort study investigating the prenatal and early postnatal determinants of child health and development. 2002 pregnant women were recruited before 24 weeks of amenorrhoea in two maternity clinics from middle-sized French cities (Nancy and Poitiers). From May 2003 to September 2006, 1899 newborns were then included. The main outcomes of interest are fetal (via ultra-sound) and postnatal growth, adiposity development, respiratory health, atopy, behaviour and bone, cognitive and motor development. We are studying fetal weight that depends on the gestational age in the second and the third trimesters of mother’s pregnancy. Some classical empirical and parametric methods as polynomials are first used to construct these curves. Polynomial regression is one of the most common parametric approach for modelling growth data especially during the prenatal period. However, some of them require strong assumptions. So, we propose to work with semi-parametric LMS method, by modifying the response variable (fetal weight) with a Box-cox transformation. Nonparametric methods as Nadaraya-Watson kernel estimation or local polynomial estimation are also proposed to construct these curves. It is the object of the communication [28] and a paper is in progress. In addition, we want to develop a test, based on Z-scores, to detect any slope breaks in the fetal development curves (work in progress).

Many articles were devoted to the problem of recursively estimating eigenvectors corresponding to eigenvalues in decreasing order of the expectation of a random matrix using an i.i.d. sample of it. The present study makes the following contributions: the convergence of processes to normed eigenvectors is proved under two sets of more general assumptions, the observed random matrices are no more supposed i.i.d.; moreover, the scope of these processes is widened. The application to online principal component analysis of a data stream is treated, assuming that data are realizations of a random vector  $Z$  whose expectation is unknown and is estimated online, as well as possibly the metric used when it depends on unknown characteristics of  $Z$ ; two types of processes are studied: we are no more bound to use a data mini-batch at each step, but we can use all previously observed data up to the current step without storing them, thus taking into account all the information contained in previous data. The conducted experiments have shown that processes of the second type are faster than those of the first type. It is the object of the submitted paper [32] and the communication [21].

The study addresses the problem of constrained binary logistic regression, particularly in the case of a data stream, using a stochastic approximation process. To avoid a numerical explosion which can be encountered, we propose to use a process with online standardized data instead of raw data. This type of process can also be used when we have to standardize the explanatory variables, for example in the case of a shrinkage method such as LASSO. Herein, we define and study the almost sure convergence of an averaged constrained stochastic gradient process with online standardized data. Moreover we propose to use a piecewise constant

step-size in order that the step-size does not decrease too quickly and reduce the speed of convergence. Processes of this type are compared to classical processes on real and simulated datasets. The results of conducted experiments confirm the validity of the choices made. This will be used in an ongoing application to online updating of a score in heart failure patients. It is the object of the submitted paper [31] and the communications [20],[19].

## 8. Bilateral Contracts and Grants with Industry

### 8.1. Bilateral Contracts with Industry

Bruno Scherrer has done some consulting for EDF. This was a skill transfer activity involving training and consulting on the theory and algorithms for reinforcement learning, for the Research & Development team of EDF lead by Lorenzo Audibert. This R&D team wants to apply reinforcement learning to several EDF problems: optimizing maintenance of uranium rods in the cores of nuclear power plants, optimization of load profiles for a network of electric vehicles. Bruno Scherrer's role was to give them the basics of reinforcement learning theory, and help them to use the algorithms of the literature. It was a one-shot action, running in 2018 and 2019, and contractualized via a "framework agreement" Inria-EDF. This contract brings in approximately 12,000 euros to BIGS team (among which 2,000 for mission expenses).

R. Azaïs, A. Gégout-Petit, F. Greciet collaborated with SAFRAN Aircraft Engines (through a 2016-2019 contract). SAFRAN Aircraft Engines designs and products aircraft engines. For the design of pieces, they have to understand the mechanism of crack propagation under different conditions. BIGS models crack propagation with Piecewise Deterministic Markov Processes (PDMP).

## 9. Partnerships and Cooperations

### 9.1. Regional Initiatives

- Lorraine Université d'Excellence LUE, Impact Project GEENAGE (Functional Genomic, Epigenomic and ENvironment interplay to impact the understanding, diagnosis and management of healthy and pathological AGEing). Anne Gégout-Petit, Lionel Lenôtre, Emma Horton.

### 9.2. National Initiatives

- FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARTERial AGEing ; leader : Pr Athanase Benetos), Jean-Marie Monnez, Benoît Lalloué, Anne Gégout-Petit.
- RHU Fight HF (Fighting Heart Failure ; leader : Pr Patrick Rossignol), located at the University Hospital of Nancy, Jean-Marie Monnez, Benoît Lalloué.
- 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
- A. Gégout-Petit, N. Sahki, S. Mézières are involved in the learning aspect of the clinical protocol "EOLEVAL" with Assistance Publique des Hopitaux de Paris (APHP)
- "ITMO Physics, mathematics applied to Cancer" (2017-2019): "Modeling ctDNA dynamics for detecting targeted therapy", Funding organisms: ITMO Cancer, ITMO Technologies pour la santé de l'alliance nationale pour les sciences de la vie et de la santé (AVIESAN), INCa, Leader: N. Champagnat (Inria TOSCA), Participants: A. Gégout-Petit, A. Muller-Gueudin, P. Vallois.
- PEPS AMIES (2019-2020), Etude Biométrique en foetopathologie et développement de l'enfant, Collaboration between Institut Elie Cartan and the CRESS INSERM, S. Ferrigno.
- Modular, multivalent and multiplexed tools for dual molecular imaging (2017-2020), Funding organism: ANR, Leader: B Kuhnast (CEA). Participant: T. Bastogne.

- Sophie Mézières belongs to GDR 720 ISIS, Funding organism: CNRS, leader: Laure Blanc-Féraud.

## 9.3. International Research Visitors

### 9.3.1. Visits of International Scientists

Juhyun Park from Bath University spent a week in Nancy in June 2019 to work on tests for paired distributions in the framework of functional analysis with Anne Gégout-Petit.

## 10. Dissemination

### 10.1. Promoting Scientific Activities

#### 10.1.1. Scientific Events: Organisation

##### 10.1.1.1. Member of the Organizing Committees

- Sandie Ferrigno, Anne Gégout-Petit, Clémence Karmann, Aurélie Muller-Gueudin, Nassim Sahki, Bruno Scherrer and Sophie Wantz-Mézières were very active members of the organizing committee of the "Journées de Statistique de la SFdS" at FST in Nancy, June 2019.
- Anne Gégout-Petit, Aurélie Muller-Gueudin, and Pierre Vallois were in the organizing committee of the workshop "Modélisation de l'hétérogénéité tumorale et thérapies ciblées" at FST in Nancy, September 2019.
- Sophie Wantz-Mézières was part of the organizing committee of the national workshop of the "Fédération Charles Hermite": "Sécurité et confiance dans les échanges des données de santé", at the "Ecole de Chirurgie" in Nancy, November 2019.

#### 10.1.2. Scientific Events: Selection

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

A. Gégout-Petit was ETC program coordinator of the European Meeting of Statisticians, Palermo, July 2019.

##### 10.1.2.2. Member of the Conference Program Committees

A. Gégout-Petit was part of the program committee of the CFIES (Colloque francophone international sur l'enseignement de la statistique), Strasbourg, Septembre 2019.

#### 10.1.3. Journal

##### 10.1.3.1. Reviewer - Reviewing Activities

All BIGS members are regular reviewers for journals in probability, statistics and machine learning: 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

- Talk "Agrégation de méthodes statistiques pour la sélection de variables corrélées et en grande dimension". Séminaire AgroParisTech, Paris, May 2019, Aurélie Muller-Gueudin.
- Talk "Modélisation de réseaux de régulation de gènes par processus déterministes par morceaux". Journée de la Fédération Charles Hermite, Nancy, June 2019, Aurélie Muller-Gueudin.

#### 10.1.5. Research Administration

Anne Gégout-Petit is member of "Bureau du comité des projets", centre Inria Nancy Grand-Est.



## 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching

B. Scherrer excepted, BIGS members have teaching obligations at "Université Lorraine" and are teaching at least 192 hours each year. They teach probability and statistics at different levels (Licence, Master, Engineering school). Many of them have pedagogical responsibilities.

- 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
- T. Bastogne is in charge of the spécialité Systèmes & TIC du master Ingénierie de Systèmes Complexes
- T. Bastogne is in charge of professional master: CIIBLE (Cybernétique, Instrumentation, Image en Biologie et medecinE) en M2 with Medicine Faculty of Université de Lorraine
- T. Bastogne is 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: 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)
- 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

### 10.2.2. Supervision

- PhD : Clémence Karmann, "Network inference for zero-inflated models", Grant : Inria-Cordis. Advisors: A. Gégout-Petit, A. Muller-Gueudin, Université de Lorraine, defended on November 25, 2019.
- PhD : Florine Greciet, "Modèles markoviens déterministes par morceaux cachés pour la propagation de fissures", grant CIFRE SAFRAN AIRCRAFT ENGINES, Advisors : R. Azaïs, A. Gégout-Petit, Université de Lorraine, defense on January, 2020.
- PhD : Kévin Duarte, "Aide à la décision médicale et télémédecine dans le suivi de l'insuffisance cardiaque", Advisors : J.-M. Monnez and E. Albuisson.
- PhD : Pauline Guyot, "Modélisation et Simulation de l'Electrocardiogramme d'un Patient Numérique", Grant : CIFRE-Cybernano. Advisors: T. Bastogne, E. H. Djermoune.
- PhD: Nassim Shaki, "Détection de rupture dans des signaux multivariés pour la prédiction d'événement redouté à partir de paramètres physiologiques recueillis par capteurs connectés après greffe pulmonaire", grant Inria-Cordis. Advisors: A. Gégout-Petit, S. Mézières, M. d'Ortho.
- Benoît Lalloué, contract research engineer for two years, RHU Fight RF, supervised by Jean-Marie Monnez.
- Postdoc: Lionel Lenôtre, Telomer Modelling, grant LUE GEENAGE. Advisors: A. Gégout-Petit, D. Villemonais.
- Postdoc: Emma Horton, Telomer Modelling, grant LUE GEENAGE. Advisors: A. Gégout-Petit, D. Villemonais.
- Master: all BIGS members regularly supervise project and internship of master IMOI students.
- Engineering school: all BIGS members regularly supervise project of "Ecole des Mines ", ENSEM or EEIGM students.

### 10.2.3. Juries

- Anne Gégout-Petit participated to the jury of the Phd defense of Edouard Fournier, Paul Sabatier University, Toulouse, October 8th
- Anne Gégout-Petit participated to the jury of the HDR defense of Denis Villemonais, Lorraine Université, November 18th
- Anne Gégout-Petit was reviewer and participated to the jury of the Phd defense of Candy Abboud, INRA d'Avignon, Université d'Aix-Marseille, December 11th.
- Anne Gégout-Petit participated to the jury of the HDR defense of Adrien Coulet, Lorraine Université, December 16th.
- Anne Gégout-Petit is member of the "Jury du prix de thèse AMIES".

## 10.3. Popularization

### 10.3.1. Interventions

- Anne Gégout-Petit participated in a round table on Artificial Intelligence and Health as part of "La fête de la science" 2019
- Anne Gégout-Petit participated at the "Brunch Données de santé et thermalisme" Janvier 2019
- Sandie Ferrigno: Advisor of a group of students (EEIGM), "La main à la Pâte" project, elementary schools, Nancy, January-June 2019
- Sandie Ferrigno: Advisor of a group of students (EEIGM), "Energies renouvelables", "La main à la Pâte" project, Institut médico-éducatif (IME), Commercy, October 2019-January 2020
- Sandie Ferrigno: Advisor of a group of students (EEIGM), "L'Astronomie" Cgénial project, Collège Paul Verlaine, Malzéville, November 2019-February 2020.

- Sandie Ferrigno: Advisor of a group of students (EEIGM), "Le Chocolat" Cgénial project, Collège de la Craffe, Nancy, November 2019-January 2020.
- Sophie Wantz-Mézières was part of the organization of a thematic and multidisciplinary week "Neurosciences, neuro-oncologie et numérique", for students from Telecom-Nancy and Faculté de Médecine de Nancy, May 2019.

## 11. Bibliography

### Publications of the year

#### Doctoral Dissertations and Habilitation Theses

- [1] C. KARMANN. *Network inference for zero-inflated models*, Université de Lorraine (Nancy), November 2019, <https://hal.archives-ouvertes.fr/tel-02384511>

#### Articles in International Peer-Reviewed Journals

- [2] J.-B. BARBRY, A.-S. POINSARD, T. BASTOGNE, O. BALLAND. *Short-term effects of ocular 2% dorzolamide, 0.5% timolol or 0.005% latanoprost on the anterior segment architecture in healthy cats: a prospective study*, in "Open Veterinary Journal", 2020, forthcoming, <https://hal.archives-ouvertes.fr/hal-02396549>
- [3] M. BEN ABDALLAH, M. BLONSKI, S. WANTZ-MÉZIÈRES, Y. GAUDEAU, L. TAILLANDIER, J.-M. MOUREAUX, A. DARLIX, N. MENJOT DE CHAMPFLEUR, H. DUFFAU. *Data-driven predictive models of diffuse low-grade gliomas under chemotherapy*, in "IEEE Journal of Biomedical and Health Informatics", January 2019, vol. 23, n<sup>o</sup> 1, pp. 38-46 [DOI : 10.1109/JBHI.2018.2834159], <https://hal.archives-ouvertes.fr/hal-02097695>
- [4] A. GÉGOUT-PETIT, L. GUÉRIN-DUBRANA, S. LI. *A new centered spatio-temporal autologistic regression model with an application to local spread of plant diseases*, in "Spatial Statistics", May 2019, n<sup>o</sup> 31, <https://arxiv.org/abs/1811.06782>, <https://hal.inria.fr/hal-01926115>
- [5] A. LEJAY, L. LENÔTRE, G. PICHOT. *An exponential timestepping algorithm for diffusion with discontinuous coefficients*, in "Journal of Computational Physics", November 2019, vol. 396, pp. 888-904 [DOI : 10.1016/J.JCP.2019.07.013], <https://hal.inria.fr/hal-01806465>
- [6] F. RECH, G. HERBET, Y. GAUDEAU, S. WANTZ-MÉZIÈRES, J.-M. MOUREAUX, S. MORITZ-GASSER, H. DUFFAU. *A probabilistic map of negative motor areas of the upper limb and face: a brain stimulation study*, in "Brain - A Journal of Neurology", April 2019, vol. 142, n<sup>o</sup> 4, pp. 952-965 [DOI : 10.1093/BRAIN/AWZ021], <https://hal.archives-ouvertes.fr/hal-02314548>
- [7] A. TACCOEN, C. C. PIEDALLU, I. SEYNAVE, V. V. PEREZ, A. GÉGOUT-PETIT, L.-M. NAGELEISEN, J.-D. BONTEMPS, J.-C. GÉGOUT. *Background mortality drivers of European tree species: climate change matters*, in "Proceedings of the Royal Society B: Biological Sciences", April 2019, vol. 286, n<sup>o</sup> 1900, 20190386 p. [DOI : 10.1098/RSPB.2019.0386], <https://hal.archives-ouvertes.fr/hal-02095574>
- [8] S. TOUPANCE, D. VILLEMONAIS, D. GERMAIN, A. GÉGOUT-PETIT, E. ALBUISSON, A. BENETOS. *The individual's signature of telomere length distribution*, in "Scientific Reports", January 2019, vol. 9, n<sup>o</sup> 1, 8 p. [DOI : 10.1038/s41598-018-36756-8], <https://hal.inria.fr/hal-01925000>

## International Conferences with Proceedings

### [9] *Best Paper*

Y. EFRONI, G. DALAL, B. SCHERRER, S. MANNOR. *How to Combine Tree-Search Methods in Reinforcement Learning*, in "AAAI 19 - Thirty-Third AAAI Conference on Artificial Intelligence", Honolulu, Hawaii, United States, January 2019, <https://arxiv.org/abs/1809.01843> - AAAI 2019, <https://hal.inria.fr/hal-02273713>.

[10] M. GEIST, B. SCHERRER, O. PIETQUIN. *A Theory of Regularized Markov Decision Processes*, in "ICML 2019 - Thirty-sixth International Conference on Machine Learning", Long Island, United States, June 2019, <https://arxiv.org/abs/1901.11275> - ICML 2019, <https://hal.inria.fr/hal-02273741>

[11] N. SAHKI, A. GÉGOUT-PETIT, S. WANTZ-MÉZIÈRES. *Détection Statistique de Rupture dans le Cadre Online*, in "JdS 2019 - 51èmes Journées de Statistique", Nancy, France, June 2019, <https://hal.inria.fr/hal-02289680>

## Conferences without Proceedings

[12] B. BASTIEN, A. GÉGOUT-PETIT, A. MULLER-GUEUDIN. *Aggregation of statistical methods for the selection of correlated and high dimensional variables*, in "Séminaire AgroParisTech", Paris, France, May 2019, <https://hal.archives-ouvertes.fr/hal-02360968>

[13] J. DELEFORTERIE, Y. KOLASA, L. BATISTA, J. HUTIN, T. BASTOGNE. *Quality-by-Design for the safe development of medical devices containing nanomaterials. A study case in photodynamic therapy*, in "NanoMed Europe, NME 2019", Braga, Portugal, June 2019, Présentation Poster, <https://hal.archives-ouvertes.fr/hal-02396520>

[14] F. GRECIET, R. AZAÏS, A. GÉGOUT-PETIT. *Régression polynomiale par morceaux sous contrainte de régularité pour la propagation de fissures*, in "JdS 2019 - 51èmes Journées de Statistique", Nancy, France, June 2019, pp. 1-5, <https://hal.archives-ouvertes.fr/hal-02172747>

[15] A. GÉGOUT-PETIT, A. GUEUDIN, C. KARMANN. *Network inference for truncated gaussian data*, in "European Meeting of Statisticians", Palermo, Italy, July 2019, <https://hal.archives-ouvertes.fr/hal-02369239>

[16] C. KARMANN, A. GÉGOUT-PETIT, A. MULLER-GUEUDIN. *Inférence de réseaux pour des données gaussiennes inflatées en zéros par double troncature*, in "Journées de statistique 2019", Nancy, France, June 2019, <https://hal.archives-ouvertes.fr/hal-02335105>

[17] C. KARMANN, A. GÉGOUT-PETIT, A. MULLER-GUEUDIN. *Méthode des knockoffs revisités pour la sélection de variables. Application à l'inférence de réseaux pour modèles inflatés en zéro*, in "Journées NETBIO Saclay", Saclay, France, October 2019, <https://hal.archives-ouvertes.fr/hal-02354748>

[18] C. KARMANN, A. GÉGOUT-PETIT, A. MULLER-GUEUDIN. *Penalized ordinal logistic regression using cumulative logits*, in "Journée scientifique FCH : "Méthodes et modèles pour comprendre les réseaux biologiques"", Nancy, France, 2019, <https://hal.archives-ouvertes.fr/hal-02354731>

[19] B. LALLOUÉ, J.-M. MONNEZ, E. ALBUISSON. *Actualisation en ligne d'un score d'ensemble*, in "51e Journées de Statistique", Nancy, France, Société Française de Statistique, June 2019, <https://hal.archives-ouvertes.fr/hal-02152352>

- [20] B. LALLOUÉ, J.-M. MONNEZ, E. ALBUISSON. *Streaming constrained binary logistic regression with online standardized data*, in "SFC 2019 - 26èmes Rencontres de la Société Francophone de Classification", Nancy, France, September 2019, <https://hal.archives-ouvertes.fr/hal-02278090>
- [21] J.-M. MONNEZ. *Convergence du processus de Oja et ACP en ligne*, in "51èmes Journées de Statistique, Nancy, 2019", NANCY, France, Efoevi Koudou, June 2019, <https://hal.archives-ouvertes.fr/hal-02383570>
- [22] A. MULLER-GUEUDIN, A. DEBUSSCHE, A. CRUDU. *Modeling of gene regulation networks by deterministic processes by pieces*, in "Journée de la Fédération Charles Hermite", Vandoeuvre-les-Nancy, France, June 2019, pp. 1-37, <https://hal.archives-ouvertes.fr/hal-02360992>
- [23] A. MULLER-GUEUDIN, A. GÉGOUT-PETIT. *Aggregation of statistical methods for the selection of correlated variables, in large dimension*, in "JdS 2019 - 51emes Journées de Statistique de la SFDS", Vandoeuvre-les-Nancy, France, June 2019, <https://hal.archives-ouvertes.fr/hal-02360974>
- [24] R. POSTOYAN, M. GRANZOTTO, L. BUŞONIU, B. SCHERRER, D. NEŠIĆ, J. DAAFOUZ. *Stability guarantees for nonlinear discrete-time systems controlled by approximate value iteration*, in "58th IEEE Conference on Decision and Control, CDC 2019", Nice, France, December 2019, Version longue de l'article du même titre et des mêmes auteurs des proceedings de l'IEEE Conference on Decision on Control 2019, Nice, France, <https://hal.archives-ouvertes.fr/hal-02271268>
- [25] V. ROULETTE, G. DELPLANQUE, J. DELEFORTERIE, T. BASTOGNE. *A contribution in nanoinformatics to facilitate the collection of structured data for Quality-by-Design in nanomedicine*, in "NanoMed Europe, NME 2019", Braga, Portugal, June 2019, <https://hal.archives-ouvertes.fr/hal-02396540>
- [26] N. SAHKI, A. GÉGOUT-PETIT, S. WANTZ-MÉZIÈRES. *New Detection Thresholds and Stop Rules for CUSUM Online Detection*, in "ENBIS 2019 - 19th Annual Conference of the European Network for Business and Industrial Statistics", Budapest, Hungary, September 2019, <https://hal.archives-ouvertes.fr/hal-02289501>

### Other Publications

- [27] B. BASTIEN, T. BOUKHOBZA, H. DUMOND, A. GÉGOUT-PETIT, A. MULLER-GUEUDIN, C. THIÉBAUT. *A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology*, September 2019, <https://arxiv.org/abs/1909.05481> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02173568>
- [28] S. FERRIGNO, M. MAUMY-BERTRAND. *Estimation of reference curves for fetal weight*, December 2019, CMStatistics 2019, Poster, <https://hal.inria.fr/hal-02389157>
- [29] A. GÉGOUT-PETIT, A. MULLER-GUEUDIN, C. KARMANN. *Graph estimation for Gaussian data zero-inflated by double truncation*, November 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02367344>
- [30] A. GÉGOUT-PETIT, A. MULLER-GUEUDIN, C. KARMANN. *The revisited knockoffs method for variable selection in  $L_1$ -penalised regressions*, November 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01799914>

- [31] B. LALLOUÉ, J.-M. MONNEZ, E. ALBUISSON. *Streaming constrained binary logistic regression with online standardized data. Application to scoring heart failure*, June 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-02156324>
- [32] J.-M. MONNEZ, A. SKIREDJ. *Convergence of a normed eigenvector stochastic approximation process and application to online principal component analysis of a data stream*, May 2019, working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01844419>
- [33] A. MULLER-GUEUDIN, A. GÉGOUT-PETIT. *Package 'armada' : A Statistical Methodology to Select Covariates in High-Dimensional Data under Dependence*, April 2019, An R package, available on the CRAN. A Statistical Methodology to Select Covariates in High-Dimensional Data under Dependence, <https://hal.archives-ouvertes.fr/hal-02363338>
- [34] N. SAHKI, A. GÉGOUT-PETIT, S. WANTZ-MÉZIÈRES. *Change-point detection method for the prediction of dreaded events during online monitoring of lung transplant patients*, December 2019, Annual PhD students conference IAEM Lorraine, APIL 2019, Poster, <https://hal.inria.fr/hal-02392756>
- [35] N. SAHKI, A. GÉGOUT-PETIT, S. WANTZ-MÉZIÈRES. *Performance Study of Detection Thresholds for CUSUM statistic in a Sequential Context*, December 2019, working paper or preprint, <https://hal.inria.fr/hal-02389331>

## References in notes

- [36] 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
- [37] 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
- [38] R. AZAÏS, F. DUFOUR, A. GÉGOUT-PETIT. *Non-Parametric Estimation of the Conditional Distribution of the Interjumping Times for Piecewise-Deterministic Markov Processes*, in "Scandinavian Journal of Statistics", December 2014, vol. 41, n<sup>o</sup> 4, pp. 950–969 [DOI : 10.1111/SJOS.12076], <https://hal.archives-ouvertes.fr/hal-01103700>
- [39] 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>
- [40] J. M. BARDET, G. LANG, G. OPPENHEIM, A. PHILIPPE, S. STOEVEV, M. TAQUU. *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-MÉZIÈRES, 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] D. BERTSEKAS, J. TSITSIKLIS. *Neurodynamic Programming*, Athena Scientific, 1996
- [44] 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/>
- [45] H. CARDOT, P. CÉNAC, J.-M. MONNEZ. *A fast and recursive algorithm for clustering large datasets with k-medians*, in "Computational Statistics & Data Analysis", 2012, vol. 56, n<sup>o</sup> 6, pp. 1434–1449
- [46] 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
- [47] 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
- [48] 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>
- [49] 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
- [50] 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
- [51] S. FERRIGNO, M. MAUMY-BERTRAND, A. MULLER-GUEUDIN. *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>
- [52] 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
- [53] 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
- [54] T. HANSEN, U. ZWICK. *Lower Bounds for Howard's Algorithm for Finding Minimum Mean-Cost Cycles*, in "ISAAC (1)", 2010, pp. 415-426
- [55] 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>
- [56] 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

- [57] 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>
- [58] R. KOENKER. *Quantile regression*, Cambridge university press, 2005, n<sup>o</sup> 38
- [59] Y. A. KUTOYANTS. *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London Ltd., London, 2004, xiv+481 p.
- [60] C. LACAUX. *Real Harmonizable Multifractional Lévy Motions*, in "Ann. Inst. Poincaré.", 2004, vol. 40, n<sup>o</sup> 3, pp. 259–277
- [61] 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
- [62] B. LESNER, B. SCHERRER. *Non-Stationary Approximate Modified Policy Iteration*, in "ICML 2015", Lille, France, July 2015, <https://hal.inria.fr/hal-01186664>
- [63] T. LYONS, Z. QIAN. *System control and rough paths*, Oxford mathematical monographs, Clarendon Press, 2002, <http://books.google.com/books?id=H9fRQNIngZYC>
- [64] N. MEINSHAUSEN, P. BÜHLMANN. *High-dimensional graphs and variable selection with the lasso*, in "The Annals of Statistics", 2006, pp. 1436–1462
- [65] 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
- [66] 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>
- [67] 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
- [68] 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
- [69] I. POST, Y. YE. *The simplex method is strongly polynomial for deterministic Markov decision processes*, arXiv:1208.5083v2, 2012
- [70] M. PUTERMAN. *Markov Decision Processes*, Wiley, New York, 1994
- [71] B. ROYNETTE, P. VALLOIS, M. YOR. *Brownian penalisations related to excursion lengths, VII*, in "Annales de l'IHP Probabilités et statistiques", 2009, vol. 45, n<sup>o</sup> 2, pp. 421–452
- [72] 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



- [73] 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)
- [74] 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>
- [75] 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>
- [76] 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>
- [77] 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>
- [78] 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>
- [79] 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>
- [80] P. VALLOIS. *The range of a simple random walk on  $\mathbb{Z}$* , in "Advances in applied probability", 1996, pp. 1014–1033
- [81] 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)
- [82] 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