Activity Report 2014

Project-Team BIGS

Biology, genetics and statistics

Table of contents

Keywords: Computational Biology, Data Analysis, Identification, Statistical Methods, Stochastic Modeling

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

1. Members

Research Scientist
  Romain Azaïs [Inria, Researcher, from Oct. 2014]

Faculty Members
  Samy Tindel [Team leader, Univ. Lorraine, Professor, HdR]
  Thierry Bastogne [Univ. Lorraine, Professor, HdR]
  Sandie Ferrigno [Univ. Lorraine, Associate Professor]
  Céline Lacaux [Univ. Lorraine, Associate Professor, HdR]
  Sophie Wantz-Mézières [Univ. Lorraine, Associate Professor]
  Jean-Marie Monnez [Univ. Lorraine, Professor, HdR]
  Aurélie Muller-Gueudin [Univ. Lorraine, Associate Professor]
  Pierre Vallois [Univ. Lorraine, Professor, HdR]

PhD Student
  Kevin Duarte [Univ. Lorraine, from Sept. 2013]

Post-Doctoral Fellows
  Laura Vinckenbosch [Inria, until Feb 2014]
  Juan-José Viquez Rodriguez [Inria, until Jan 2014]

Administrative Assistant
  Isabelle Blanchard [Inria]

2. Overall Objectives

2.1. Overall Objectives

BIGS is a team labeled by Inria, by CNRS and by Université de Lorraine, via the Institut Élie Cartan (UMR 7502 CNRS-Inria-UL). Our research is mainly focused on statistics and stochastic processes techniques aiming at a better understanding of biological systems. A special attention is devoted to online data analysis, local regression techniques and identification of complex biological systems. Our investigations encompass both theoretical aspects and concrete applications of the issues alluded to above. To be more specific, we focus on the following topics:

●*Online Factorial Analysis:* High dimensional data are often obtained online, and cannot be stored integrally in a computer memory. One of the recent challenges in data analysis is then to be able to perform an accurate classification or clustering by taking advantage of the possibility of updating the information. This has to be done, of course, in a rather simple and efficient way, allowing real time analysis. To this aim, we use techniques based on some sophisticated tools coming from stochastic approximation.

●*Local Regression Techniques:* The main issue here is the construction of a procedure allowing to assess, in quite a general framework, whether a given model fits a data set regarding most assumptions made in elaborating the model. This is based on a generalization of the Cramér-Von Mises statistics and involves a non parametric estimate of the conditional distribution of the response variable. A detailed analysis of the procedure, including rate of convergence and asymptotic properties, is being performed. The strategy is then implemented for a study concerning fetal biometry.

•*Photodynamic therapy:* Since 1988, some control system scientists and biologists at the Centre de Recherche en Automatique de Nancy (CRAN in short) 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.

•*Estimation for complex biological systems:* Numerous biological systems are accurately described by multi-dimensional noisy differential equations driven by Gaussian processes (beyond the realm of Brownian motion) or by fractional fields, for which asymptotic properties and parameter estimation are fruitful informations. We are thus interested in studying this kind of systems, having in mind 3 specific applications of interest for us: Bacteriophage systems, random fluctuation of nanoparticles and automatic detection of osteoporosis.

3. Research Program

3.1. Online data analysis

Participants: J.-M. Monnez, P. Vallois. Generally speaking, there exists an overwhelming amount of articles dealing with the analysis of high dimensional data. Indeed, this is one of the major challenges in statistics to-day, motivated by internet or biostatistics applications. Within this global picture, the problem of classification or dimension reduction of online data can be traced back at least to a seminal paper by Mac Queen [53], in which the k-means algorithm is introduced. This popular algorithm, constructed for classification purposes, consists in a stepwise updating of the centers of some classes according to a stream of data entering into the system. The literature on the topic has been growing then rapidly since the beginning of the 90's.

Our point of view on the topic relies on the so-called *French data analysis school*, and more specifically on Factorial Analysis tools. In this context, it was then rapidly seen that stochastic approximation was an essential tool (see Lebart's paper [50]), which allows one to approximate eigenvectors in a stepwise manner. A systematic study of Principal Component and Factorial Analysis has then been lead by Monnez in the series of papers [56], [54], [55], in which many aspects of convergences of online processes are analyzed thanks to the stochastic approximation techniques.

3.2. Local regression techniques

Participants: S. Ferrigno, A. Muller-Gueudin. In the context where a response variable Y is to be related to a set of regressors X, one of the general goals of Statistics is to provide the end user with a model which turns out to be useful in predicting Y for various values of X. Except for the simplest situations, the determination of a good model involves many steps. For example, for the task of predicting the value of Y as a function of the covariate X, statisticians have elaborated models such as the regression model with random regressors:

$$Y = g(X, \theta) + \sigma(X)\epsilon.$$

Many assumptions must be made to reach it as a possible model. Some require much thinking, as for example, those related to the functional form of $g(\cdot, \theta)$. Some are made more casually, as often those related to the functional form of $\sigma(\cdot)$ or those concerning the distribution of the random error term ϵ. Finally, some assumptions are made for commodity. Thus the need for methods that can assess if a model is concordant with the data it is supposed to adjust. The methods fall under the banner of goodness of fit tests. Most existing tests are *directional*, in the sense that they can detect departures from only one or a few aspects of a null model. For example, many tests have been proposed in the literature to assess the validity of an entertained structural part $g(\cdot, \theta)$. Some authors have also proposed tests about the variance term $\sigma(\cdot)$ (cf. [51]). Procedures testing the normality of the ϵ_i are given, but for other assumptions much less work has been done. Therefore the need of a global test which can evaluate the validity of a global structure emerges quite naturally.

With these preliminaries in mind, let us observe that one quantity which embodies all the information about the joint behavior of (X, Y) is the cumulative conditional distribution function, defined by

$$F(y|x) = P(Y \le y | X = x).$$

The (nonparametric) estimation of this function is thus of primary importance. To this aim, notice that modern estimators are usually based on the local polynomial approach, which has been recognized as superior to classical estimates based on the Nadaraya-Watson approach, and are as good as the recent versions based on spline and other methods. In some recent works [41], [42], we address the following questions:

- Construction of a global test by means of Cramér-von Mises statistic.
- Optimal bandwidth of the kernel used for approximation purposes.

We also obtain sharp estimates on the conditional distribution function in [43].

3.3. Stochastic modeling for complex and biological systems

Participants: R. Azaïs, T. Bastogne, C. Lacaux, A. Muller-Gueudin, S. Tindel, P. Vallois, S. Wantz-Mézières

In most biological contexts, mathematics turn out to be useful in producing accurate models with dual objectives: they should be simple enough and meaningful for the biologist on the one hand, and they should provide some insight on the biological phenomenon at stake on the other hand. We have focused on this kind of issue in various contexts that we shall summarize below.

Photodynamic Therapy: Photodynamic therapy induces a huge demand of interconnected mathematical systems, among which we have studied recently the following ones:

• The tumor growth model is of crucial importance in order to understand the behavior of the whole therapy. We have considered the tumor growth as a stochastic equation, for which we have handled the problem uncertainties on the measure times [27] as well as mixed effects for parameter estimation.

• Another important aspect to quantify for photodynamic therapy calibration is the response to radiotherapy treatments. There are several valid mathematical ways to describe this process, among which we distinguish the so-called hit model. This model assumes that whenever a group of sensitive targets (chromosomes, membrane) in the cell are reached by a sufficient number of radiations, then the cell is inactivated and dies. We have elaborated on this scheme in order to take into account two additional facts: (i) The reduction of the cell situation to a two-state model might be an oversimplification. (ii) Several doses of radiations are inoculated as time passes. These observations have lead us to introduce a new model based on multi-state Markov chains arguments (Keinj & al, 2012), in which cell proliferation can be incorporated.

Bacteriophage therapy: Let us mention a starting collaboration between BIGS and the Genetics and Microbiology department at the Universitat Autònoma de Barcelona, on the modeling of bacteriophage therapies. The main objective here is to describe how a certain family of benign viruses is able to weaken a bacterium induced disease, which naturally leads to the introduction of a noisy predator-prey system of equations. It should be mentioned that some similar problems have been treated (in a rather informal way, invoking a linearization procedure) by Carletti in [34]. These tools cannot be applied directly to our system, and our methods are based on concentration and large deviations techniques (on which we already had an expertise [57], [60]) in order to combine convergence to equilibrium for the deterministic system and deviations of the stochastic system. Notice that A. Muller-Gueudin is also working with A. Debussche and O. Radulescu on a related topic [37], namely the convergence of a model of cellular biochemical reactions.

Gaussian signals: 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 differential equations driven by a fractional Brownian motion (fBm) and other Gaussian processes are now well understood, mainly thanks to the so-called *rough paths* tools [52], but also invoking the Russo-Vallois integration techniques [59]. The specific issue of Volterra equations driven by fBm, which is central for the subdiffusion within proteins problem, is addressed in [38].

Fractional fields are very often used to model irregular phenomena which exhibit a scale invariance property, fractional Brownian motion being the historical fractional model. Nevertheless, its isotropy property is a serious drawback for instance in hydrology or in medecine (see [33]). Moreover, the fractional Brownian motion cannot be used to model some phenomena for which the regularity varies with time. Hence, many generalizations (Gaussian or not) of this model have been recently proposed, see for instance [28] for some Gaussian locally self-similar fields, [46] for some non-Gaussian models, [31] for anisotropic models.

Our team has thus contributed [36], [47], [46], [48], [58] and still contributes [30], [32], [31], [49], [44] to this theoretical study: Hölder continuity, fractal dimensions, existence and uniqueness results for differential equations, study of the laws to quote a few examples. As we shall see below, this line of investigation also has some impact in terms of applications: we shall discuss how we plan to apply our results to osteoporosis on the one hand and to fluctuations within protein molecules on the other hand.

3.4. Parameter identifiability and estimation

Participants: R. Azaïs, T. Bastogne, S. Tindel, P. Vallois, S. Wantz-Mézières

When one desires to confront theoretical probabilistic models with real data, statistical tools are obviously crucial. We have focused on two of them: parameter identifiability and parameter estimation.

Parameter identifiability [62] deals with the possibility to give a unique value to each parameter of a mathematical model structure in inverse problems. There are many methods for testing models for identifiability: Laplace transform, similarity transform, Taylor series, local state isomorphism or elimination theory. Most of the current approaches are devoted to *a priori* identifiability and are based on algebraic techniques. We are particularly concerned with *a posteriori* identifiability, *i.e.*, after experiments or in a constrained experimental framework and the link with experimental design techniques. Our approach is based on statistical techniques through the use of variance-based methods. These techniques are strongly connected with global sensitivity approaches and Monte Carlo methods.

The parameter estimation for a family of probability laws has a very long story in statistics, and we refer to [29] for an elegant overview of the topic. Moving to the references more closely related to our specific projects, let us recall first that the mathematical description of photodynamic therapy can be split up into three parametric models : the uptake model (pharmacokinetics of the photosensitizing drug into cancer cells), the photoreaction model and the tumor growth model. Several papers have been reported for the application of system identification techniques to pharmacokinetics modeling problems. But two issues were ignored in these previous works: presence of timing noise and identification from longitudinal data. In [27], 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 can be performed by the *Monolix* software (http://www.lixoft.eu/products/monolix/product-monolix-overview/) developed by the Monolix group chaired by Marc Lavielle and France Mentré, and supported by Inria. In the recent past, we have used this tool for tumor growth modeling. According to what we know so far, no parameter estimation study has been reported about the photoreaction model in photodynamic therapy. A photoreaction model, composed of six stochastic differential equations, is proposed in [39]. The main open problem is to access to data. We currently build on an experimental platform which aims at overcoming this technical issue. Moreover, an identifiability study coupled to a global sensitivity analysis of the photoreaction model are currently in progress. Tumor growth is generally described by population dynamics models or by cell cycle models. Faced with this wide variety of descriptions, one of the main open problems is to identify the suitable model structure. As mentioned above, we currently investigate alternative representations based on branching processes and Markov chains, with a model selection procedure in mind.

A few words should be said about the existing literature on statistical inference for diffusion or related processes, a topic which will be at the heart of three of our projects (namely photodynamic and bacteriophage therapies, as well as fluctuations within molecules). The monograph [45] is a good reference on the basic estimation techniques for diffusion processes. The problem of estimating diffusions observed at discrete times, of crucial importance for applications, has been addressed mainly since the mid 90s. The maximum likelihood techniques, which are also classical for parameter estimation, are well represented by the contributions [40].

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 [26], [35]. On the other hand, the inference problem for diffusions driven by a fractional Brownian motion is still in its infancy. A good reference on the question is [61], dealing with some very particular families of equations, which do not cover the cases of interest for us.

4. New Software and Platforms

4.1. Online data analysis

Participants: J.-M. Monnez

An R package performing most of the methods of factorial analysis in an online way has been developed by R. Bar and J.-M. Monnez. Starting from a simulated data flow, the main goal of the program is to perform online factorial analyses (Principal Component Analyses, Canonical Correlation Analysis, Canonical Discriminant Analysis, Correspondence Analysis). Data are supposed to be independent and identically distributed observations of a random vector (whose distribution is a priori unknown). Defining stochastic approximation processes, the procedure is adaptative in the sense that the results of the analyses are updated recursively each time that a new piece of data is taken into account.

From a theoretical point of view, the i.i.d case has been recently extended to the case of an expectation and/or covariance matrix of the random vector varying with time. We plan to include these improvements into our software.

4.2. Socio-economic index

Participants: J.-M. Monnez

A R package called SesIndexCreatoR has been written by B. Lalloué and J.-M. Monnez in order to implement our socio-economic index for health inequalities. The version 1.0 of this package is currently freely available on the website of the Equit'Area project: http://www.equitarea.org/documents/packages_1.0-0/. It contains the functions needed to run the procedure (either integrally or partially) and obtain the corresponding SES index. The user may also create categories of this index with different methods (hierarchical clustering with or without k-nearest neighbors, quantiles, or intervals) and generate automatic reports of the results. Visualization and plotting functions are provided in the package.

4.3. Angio-Analytics

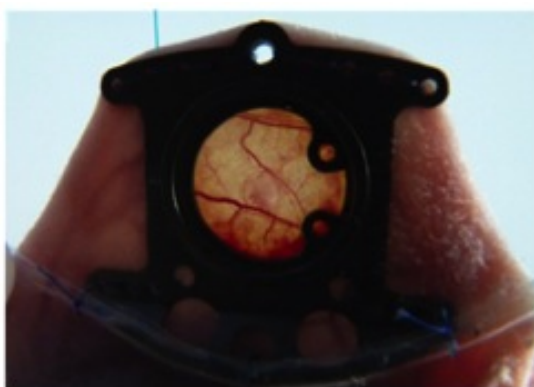*Participants: T. Bastogne*

Figure 1. Example of a skinfold chamber placed on a mouse skin

A software *Angio-Analytics* has been developed by J.-B. Tylcz, E. Djermoune and T. Bastogne. 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, as illustrated in Fig. 1. 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, as illustrated in Fig. 2. The last step is the system identification of the pharmacodynamic response and the statistical analysis of the model parameters as shown in Fig. 3 and Fig. 4. An article has been submitted to a journal (Biomedical Signal Processing and Control) and is currently in revision process. Moreover, the current version of the software has been registered to the *Agence de Protection des Programmes*.

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

Participants: T. Bastogne

More than eight million people die from cancer worldwide each year. Current treatment such as chemotherapy and radiotherapy are still limited in terms of benefit/risk ratio. Nevertheless, engineered nanoparticles have opened new interesting perspectives in cancerology, as emphasized by Brigger et al. since 2002. One of these promising solutions is based on the development of nanoparticles able to enhance the cytotoxic effect of radio-therapy. Nevertheless, the preclinical development in nano-medicine is slow, risky and expensive. Recently, Etheridge et al. (2013) highlighted the fact that many of the revolutionary nano-medicine technologies antici-pated in the literature may be 20 or more years from clinical use. To speed up the preclinical development of
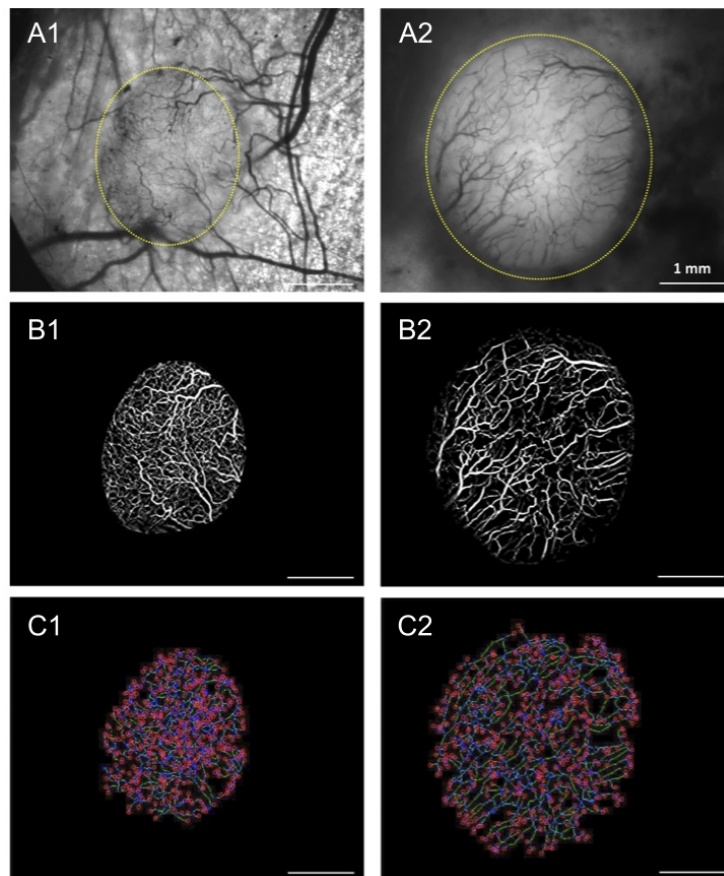
Figure 2. Example of segmentation process on a control (left) and treated tumor (right) at day -7: manual segmentation (ROI) of cancerous tissues is done in yellow on step A, vessel segmentation is performed on step B (vessels are in white), step C presents the quantification (blue and red circles) on the skeletonized vascular network (green lines)
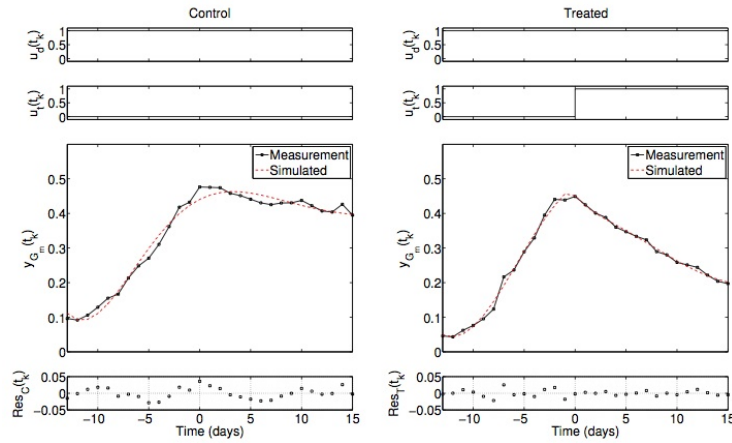
Figure 3. Measurements and estimated outputs for control and treated batches. Input signals and residuals are respectively plotted above and below

Batch	Param.	Estimate	c_v (%)
Control	b_{d_0}	0.023	12
	f_{d_1}	0.23	13
	f_{d_2}	0.057	10
	b_{t_0}	0	-
	b_{t_1}	0	-
	f_{t_1}	0	-
Treated	b_{d_0}	0.021	18
	f_{d_1}	0.16	35
	f_{d_2}	0.051	19
	b_{t_0}	−0.032	45
	b_{t_1}	−0.068	22
	f_{t_1}	0.35	50

Figure 4. Parameter estimates and coefficients of variation c_v for control and treated batches

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. That innovation gathers stochastic simulations and statistical modeling to estimate the impact of each design parameter describing the nano-object. That allows us to optimize composition factors in order to suggest one or few promising architectures regarding the medical purpose. 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. Other applications related to nano-medicines will be subject to further developments (e.g., photodynamic therapy). That contribution has received the best innovation award from the Institut Mines-Telecom in 2014 and application results will be presented at the 36th PAMM-EORTC Winter Meeting in January 2015.

5. New Results

5.1. Analysis of high dimensional data

Participants: K. Duarte, S. Ferrigno, J.-M. Monnez, A. Muller-Gueudin, S. Tindel.

5.1.1. Online partial principal component analysis of a data stream

Consider a data stream and suppose that each data vector is a realization of a random vector whose expectation varies with time, the law of the centered data vector being stationary. Consider the principal component analysis (PCA) of this centered vector called partial PCA. In this study are defined online estimations of the first principal axes by stochastic approximation processes using a data batch at each step of the process or all the data until the current step. This extends a former result obtained by J.-M. Monnez by using one data vector at each step. This is applied to partial generalized canonical correlation analysis by defining a stochastic approximation process of the metric involved in this case using all the data until the current step. If the expectation of the data vector varies according to a linear model, a stochastic approximation process of the model parameters is used. All these processes can be performed in parallel. A forthcoming preprint by R. Bar and J.-M. Monnez will discuss those aspects.

5.1.2. Data analysis for cumulative exposure Index

Everyone is subject to environmental exposures from various sources, with negative health impacts (air, water and soil contamination, noise ...) or with positive effects (e.g., green space). Studies considering such complex environmental settings in a global manner are rare. In [5] we propose to use statistical factor and cluster analyses to create a composite exposure index with a data-driven approach, in view to assess the environmental burden experienced by populations. The study was carried out in the Great Lyon area (France, 1.2M inhabitants) at the census block group (BG) scale. We used as environmental indicators ambient air NO2 annual concentrations, noise levels, proximity to green spaces, to industrial plants, to polluted sites and to road traffic. Although it cannot be applied directly for risk or health effect assessment, the resulting index can help to identify hot spots of cumulative exposure, to prioritize urban policies or to compare the environmental burden across study areas in an epidemiological framework.

5.1.3. A simultaneous stepwise covariate selection

In supervised learning the number of values of a response variable to predict can be high. Also clustering them in a few clusters can be useful to perform relevant supervised classification analysis. On the other hand selecting relevant covariates is a crucial step to build robust and efficient prediction models, especially when too many covariates are available in regard to the overall sample size. As a first attempt to solve these problems, we had already devised in a previous study an algorithm that simultaneously clusters the levels of a categorical response variable in a limited number of clusters and selects forward the best covariates by alternate minimization of Wilks's Lambda. In the project carried out this year, we first extend the former version of

the algorithm to a more general framework where Wilks's Lambda can be replaced by any model selection criterion. We also turned forward selection into stepwise selection in order to remove covariates in real time if necessary. Finally an application of our algorithm to real datasets from peanut allergy studies allowed to get confirmation of some previously published results and suggested new discoveries. The possibilities of this algorithm are promising and it is hoped to be useful for many practitioners.

5.1.4. *Prognostic value of the Strauss estimated plasma*

We describe here an application oriented study lead jointly by J.-M. Monnez and a medical team under the supervision of E. Albuisson at CHU Brabois. The objective is to assess the prognostic value of estimations of volemia, or of their variations, beyond clinical examination in a post-hoc analysis of the Eplerenone Post-Acute Myocardial Infarction (AMI) Heart Failure (HF) Efficacy and Survival Study (EPHESUS). Assessing congestion post-discharge is indeed challenging but of paramount importance to optimize patient management and prevent hospital readmissions. The analysis was performed in a subset on 4957 patients with available data (within a full dataset of 6632 patients). Study endpoint was cardiovascular death and/or hospitalization for HF between month 1 and month 3 after post-AMI HF. Estimated plasma volume variations between baseline and month 1 were estimated by the Strauss formula, which includes hemoglobin and hematocrit ratios. Other potential predictors including congestion surrogates, hemodynamic and renal variables, and medical history variables were tested. An instantaneous estimation of plasma volume at month 1, ePVS M1, was defined and also tested. Multivariate analysis was performed using stepwise logistic regression and linear discriminant analysis. In HF complicating MI, congestion assessed by the Strauss formula and an instantaneous derived measurement of plasma volume displayed an added predictive value of early cardiovascular events, beyond routine clinical assessment. Trials assessing congestion management guided by this simple tool to monitor plasma volume are warranted.

5.1.5. *Non parametric estimation of the conditional cumulative distribution function*

This project fits into the global aim of improving local regression techniques. Indeed, we propose in [21] to study the local linear estimator of the conditional distribution function. Namely, having an i.i.d. sample $(X_i, Y_i)_{1 \leq i \leq n}$, we estimate the conditional distribution function $F(t|x) = \mathbb{P}(Y \leq t|X = x)$ by:

$$\widehat{F}_n^{(1)}(t, h_n|x) = \frac{\widehat{f}_{n,2}(x, h_n)\widehat{r}_{n,0}(x, t, h_n) - \widehat{f}_{n,1}(x, h_n)\widehat{r}_{n,1}(x, t, h_n)}{\widehat{f}_{n,0}(x, h_n)\widehat{f}_{n,2}(x, h_n) - \left(\widehat{f}_{n,1}(x, h_n)\right)^2} \tag{1}$$

where $^{(1)}$ denotes the order 1 of the local polynomial estimator, $\widehat{f}_{n,j}$ stands for a kernel estimator with order j of the probability density function f_X of X, $\widehat{r}_{n,j}$ estimates the distribution of the couple (X, Y) and h_n is a bandwidth parameter.

This estimator is a particular case of the local polynomial estimators. It is the local polynomial estimator of order $p = 1$. Another simpler estimator, with order $p = 0$, is well known as the Nadaraya-Watson estimator.

We are interested in showing the advantage of this estimator over the Nadaraya-Watson estimator. We show asymptotic results for our estimator (exact rate of uniform consistency), and establish also uniform asymptotic certainty bands for the conditional cumulative distribution function.

We obtain the following result under some assumptions on the cumulative distribution F, f_X, the kernel K and the bandwidth h_n,

$$\sup_{t \in \mathbb{R}} \sup_{x \in I} \sqrt{\frac{nh_n}{\log(h_n^{-1})}} \left| \widehat{F}_n^{(1)}(t, h_n|x) - \widehat{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t, h_n|x)\right) \right| \xrightarrow[n \to +\infty]{\mathbb{P}} \sigma_F(I) \tag{2}$$

where

$$\sigma_F^2(I) = \frac{||K||_2^2}{2\inf_{x \in I} f_X(x)}.$$

As corollaries of this result, we extend our results to other statistical functions, such as the quantiles and the regression function.

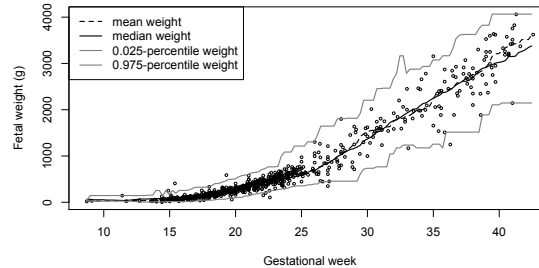We illustrate our results with simulations and an application on foetopathologic data.

Figure 5. Fetal weight during the pregnancy: estimation of mean and quantiles from our local polynomial regression method.

We have also started a study about the regression function in the application on foetopathologic data. We consider the nonparametric model

$$Y = m(X) + \epsilon,$$

where Y is the fetal weight , X are the gestational weeks, m is a smooth unknown function and ϵ the error. The goal is to provide a test to detect significant features (or change points) of this regression curve. The regression curve is estimated using local polynomial kernel smoothers.

5.2. Stochastic modeling for complex and biological systems

Participants: R. Azaïs, T. Bastogne, C. Lacaux, A. Muller-Gueudin, S. Tindel, P. Vallois, S. Wantz-Mézières

5.2.1. Modelisation of networks of multiagent systems

We relate here the beginning of collaboration between A. Gueudin, R. Azaïs and some automatic control researchers in Nancy.

We consider 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.

5.2.2. *Tumor growth modeling*

A cancer tumor can be represented for simplicity as an aggregate of cancer cells, each cell behaving according to the same discrete model and independently of the others. Therefore to measure its size evolution, it seems natural to use tools coming from dynamics of population, for instance the logistic model. This deterministic framework is well-known but the stochastic one is worthy of interest. We are currently studying in [22] a model in which we suppose that the size V_t at time t of the tumor is a diffusion process of the type :

$$\begin{cases} dV_t = r\,V_t\left(1 - \frac{V_t}{\kappa}\right) - c\,V_t + \beta\,V_t\,dB_t \\ V_0 = v > 0 \end{cases} \tag{3}$$

where $(B_t)_{t \geq 0}$ is a standard brownian motion starting from zero. Then (i) We define a family of time continuous Markov chains which models the evolution of the rate of malignant cells and approximate (under some conditions) the diffusion process (V_t). (ii) We study in depth the solution to equation (3). This diffusion process lives in a domain delimited by two boundaries: 0 and $\kappa > 0$. In this stochastic setting, the role of κ is not so clear and we contribute to understand it. We describe the asymptotic behavior of the diffusion according to the values of the parameters. The tools we resort to are boundary classification criteria and Laplace transform of the hitting time to biological worthwhile level. We are able in particular to express the mean of the hitting time.

5.2.3. *Anisotropic random fields*

Hermine Biermé (Tours) and Céline Lacaux follow in [19] their collaboration in the study of anisotropic random fields. They have extended their previous works in the framework of conditionally sub-Gaussian random series. For such anisotropic fields, they have obtained a modulus of continuity and a rate of uniform convergence. Their framework allows to study e.g., Gaussian fields, stable random fields and multi-stable random fields.

5.2.4. *Inference for dynamical systems driven by Gaussian noises.*

As mentioned in the *Scientific Foundations* Section, the problem of estimating the coefficients of a general differential equation driven by a Gaussian process is still largely unsolved. To be more specific, the most general ($\mathbb{R}$-valued) equation handled up to now as far as parameter estimation is concerned is of the form:

$$X_t^\theta = a + \theta \int_0^t b(X_u)\,du + B_t,$$

where θ is the unknown parameter, b is a smooth enough coefficient and B is a one-dimensional fractional Brownian motion. In contrast with this simple situation, our applications of interest (see the *Application Domains* Section) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\,du + \int_0^t \sigma(\theta; X_u)\,dB_t, \tag{4}$$

where θ enters non linearly in the coefficient, where σ is a non-trivial diffusion term and B is a d-dimensional fractional Brownian motion. We have thus decided to tackle this important scientific challenge first.

To this aim, here are the steps we have focused on in 2014:

- A better understanding of the underlying rough path structure for equation (4). This includes two studies on differential systems driven by some general Gaussian noises in infinite dimensions: [17] on the Parabolic Anderson model, and [16] about viscosity solutions in the rough paths setting.

- Study of densities for general systems driven by Gaussian noises as in [18] and [15].

- Ergodic aspects, which are another important ingredient for estimation procedures for stochastic differential equations, are handled in [3].

5.2.5. *Extremal process*

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 Samorodnistky have proved in [23] 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.

5.2.6. *Self-nested structure of plants*

In a recent work, Godin and Ferraro designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

The degree of self-nestedness is defined as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus design a heuristic method based on interacting simulated annealing algorithms to tackle this difficult question. This procedure is also a keystone in a new topological clustering algorithm for trees that we propose in this work. In addition, we obtain new theoretical results on the combinatorics of self-nested structures. For instance, we have shown that the number $C_{\leq H}(m)$ of self-nested trees with maximal height H and a ramification number for each vertex less than m satisfies the following formula,

$$C_{\leq H}(m) = \sum_{h=1}^{H} \prod_{i=1}^{h} \binom{m+h-i}{h-i+1}.$$

In particular, the cardinality $C_{=h}(m)$ of self-nested trees with exact height h evolves according to

$$\log C_{=h}(m) \sim \frac{(m+h)^2}{2} \log(m+h) - \frac{h^2}{2} \log h - \frac{m^2}{2} \log m - mh \log m,$$

when m and h simultaneously go to infinity. The redaction of an article is currently in progress.

5.2.7. *Semi-parametric inference for a growth-fragmentation model*

Statistical inference for piecewise-deterministic Markov processes has been extensively investigated for a few years under some ergodicity conditions. Our paper [2] is dedicated to a statistical approach for a particular non ergodic growth-fragmentation model for which the set $[0,1]$ is absorbing. This kind of stochastic process may model the dynamic of a malthusian population for which there exists an extinction threshold. We focus on the estimation of the extinction probability and of the distribution of the extinction time from only one path of the model within a long time interval.

We establish that the absorption probability p is the unique solution in an appropriate space of a Fredholm equation of the second kind whose parameters are unknown,

$$p - Kp = s,$$

where K is an integral operator depending explicitly on the main features of the model. From n data, we estimate this important characteristic of the underlying process by solving numerically the estimated Fredholm equation. Indeed, $\widehat{p}_{n,m}$ is defined as the approximated solution of the equation $p - \widehat{K}_n p = \widehat{s}_n$ after m steps of the algorithm. Fortunately, this procedure allows us to estimate also the extinction time.

We have shown the convergence in probability of the proposed estimators under some usual asymptotic conditions. In particular, we have,

$$\forall\, \varepsilon > 0, \ \mathbf{P}\left(\|p - \widehat{p}_{n,m}\|_1 > \varepsilon\right) \to 0,$$

when n and m simultaneously go to infinity. The good behavior of our estimates on finite sample sizes is presented in Figure 6. In future works, we plan to apply this methodology to more intricate situations, in particular for the pharmacokinetics and pharmacodynamics stochastic models recently introduced in the literature.
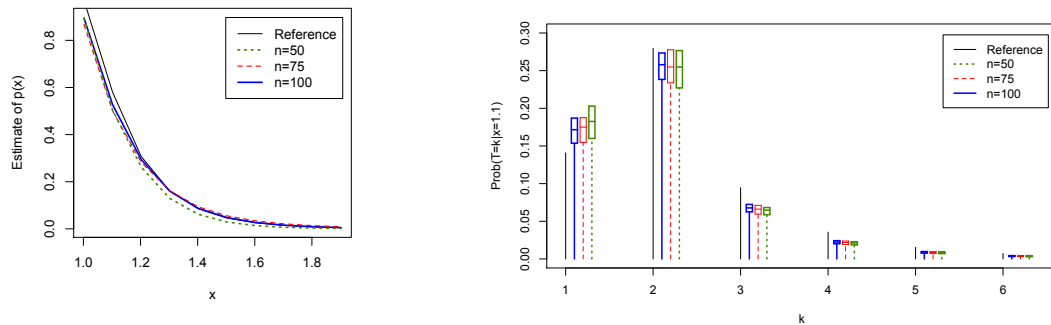

Figure 6. Probability of extinction $p(x)$ and its estimates for a trajectory starting from the initial population $1 \leq x \leq 2$ (left) and distribution of the extinction time and its estimates for a trajectory starting from $x = 1.1$ (right).

5.2.8. *A Model-based Pharmacokinetics Characterization Method of Engineered Nanoparticles for Pilot Studies*

Recent developments on engineered multifunctional nanomaterials have opened new perspectives in oncology. But assessment of both quality and safety in nanomedicine requires new methods for their biological characterization. We have recently proposed a new model-based approach for the pre-characterization of multifunctional nanomaterials pharmacokinetics in small scale in vivo studies. Two multifunctional nanoparticles, with and without active targeting, designed for photodynamic therapy guided by magnetic resonance imaging are used to exemplify the presented method. It allows the experimenter to rapidly test and select the most relevant pharmacokinetic (PK in the sequel) model structure planned to be used in the subsequent explanatory studies. We also show that the model parameters estimated from the in vivo responses provide relevant preliminary information about the tumor uptake, the elimination rate and the residual storage. For some parameters, the accuracy of the estimates is good enough to compare and draw significant pre-conclusions. A third advantage of this approach is the possibility to optimally refine the in vivo protocol for the subsequent explanatory and confirmatory studies complying with the 3Rs (reduction, refinement, replacement) ethical recommendations. More precisely, we show that the identified model may be used to select the appropriate duration of

the magnetic resonance imaging sessions planned for the subsequent studies. The proposed methodology integrates magnetic resonance image processing, continuous-time system identification algorithms and statistical analysis. Except, the choice of the model parameters to be compared and interpreted, most of the processing procedure may be automated to speed up the PK characterization process at an early stage of experimentation.

More specifically, our efforts have been split into the following tasks:

- The article [6] gives an application of statistical methods for the design of experiments to optimize the formulation of a composite molecule in photodynamic therapy. The associated know-how has been transferred to the start-up CYBERnano to be generalized to the rational design of engineered nanoparticles. Collaboration with CRAN and LRGP (Nancy) and UNINE (Neuchâtel, Suisse).

- In [12], in vivo application of photodynamic therapy, a mathematical model and computational simulations of the light propagation in biological tissues were developed to help biologists to determine *a priori* some parameters of the experimental protocol. More precisely, the numerical results were used to select the most suited position of the optical fiber to be implemented within the animal brain. This equipment is required to bring the light and thus activate the molecule within the tumor. The therapeutical objective was to maximize the homogeneity of light intensity within the tumor volume.

- Obstacles and challenges to the clinical use of the photodynamic therapy (PDT) are numerous: large inter-individual variability, heterogeneity of therapeutic predictability, lack of in vivo monitoring concerning the reactive oxygen species (ROS) production, etc. All of these factors affect in their ways the therapeutic response of the treatment and can lead to a wild uncertainty on its efficiency. To deal with these variability sources, we have designed and developed an innovative technology able to adapt in realtime the width of light impulses during the photodynamic therapy. The first objective is to accurately control the photobleaching trajectory of the photosensitizer during the treatment with a subsequent goal to improve the efficacy and reproducibility of this therapy. In this approach, the physician a priori defines the expected trajectory to be tracked by the photosensitizer photobleaching during the treatment. The photobleaching state of the PS is regularly measured during the treatment session and is used to change in real-time the illumination signal. This adaptive scheme of the photodynamic therapy has been implemented, tested and validated during in vitro tests. These tests show that controlling the photobleaching trajectory is possible, confirming the technical feasibility of such an approach to deal with inter-individual variabilities in PDT. These results, contained in [13], open new perspectives since the illumination signal can be different from a patient to another according to his individual response. This study has proven its interest by showing promising results in an in vitro context, which has to be confirmed by the current in vivo experiments. However, it is fair to say that in a near future, the proposed solution could lead, in fine, to an optimized and personalized PDT. A patent was deposited subsequently. Collaboration with CRAN (Nancy).

- The communications [8], [9] and [10] present successful applications of a model-based design of nanoparticles. This approach is based on statistical design of experiments and black-box modeling in cell biology. The associated know-how has been transferred to the start-up CYBERnano. Collaboration with CEA LETI and INSERM (Grenoble).

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

Start-up project by T. Bastogne:

- Industrial partner: CYBERnano (Contract Research Organization in NanoMedicine).
- Status: SAS created in July 2013.
- Comments: a research engineer has been hired by CYBERnano since November 2014 to develop and implement new algorithms devoted to biological signal processing.

7. Partnerships and Cooperations

7.1. Regional Initiatives

Project "Handle your heart" Creation of a drug prescription support software for the treatment of heart failure, in collaboration with the University Hospital of Nancy, headed by J.-M. Monnez.

7.2. National Initiatives

- *Truffinet* (2014), TRUFFles' microbial interaction Inference by NETwork analysis, Funding organism: PEPS CNRS-Université de Lorraine, Leader: A. Muller-Gueudin. Collaboration with IECL (Anne Gégout-Petit), CRAN (S. Martin, C. Morarescu), INRA (A. Deveau).
- *Optique-PDT* (2012-2014), mOdélisation et oPTimisation de l'Irradiance dans les tissus biologiQUEs hétérogènes traités par Thérapie PhotoDynamique interstitielle, Funding organism: *PEPS CNRS-INSERM-Inria*, Leader: M. Thomassin (CRAN, U. Lorraine).
- *Nano-Xrays* (2011-2014), Nanoparticles-based X ray-induced photodynamic therapy in glioblastoma multiform, Funding organism: Institut National du Cancer (INCa), Leader: M. Barberi-Heyob (CRAN, U. Lorraine), T. Bastogne.
- GDR 3477 Géométrie Stochastique, Leaders: Pierre Calka, David Coupier, Viet Chi Tran, C. Lacaux.
- GDR 3475 Analyse Multifractale, Leader: Stéphane Jaffard (C. Lacaux).
- *PhotoBrain* (2015-17), AGuIX® theranostic nanoparticles for vascular-targeted interstial photodynamic therapy of brain tumors, Funding organism: EuroNanoMed II, Leader: M. Barberi-Heyob (CRAN).
- (2014-16), A library of Near-InfraRed absorbing photosensitizers: tailoring and assessing photophysical and synergetic photodynamic properties, Funding organism: PHC Bosphore - Campus France, Leader: M. Barberi-Heyob (CRAN).

7.3. International Research Visitors

7.3.1. Visits of International Scientists

2014/05/11-2014/05/25: visit of Gennady Samorodnitsky (Cornell, USA) to C. Lacaux.

7.3.2. Visits to International Teams

7.3.2.1. Sabbatical programme

S. Tindel was on sabbatical at the University of Kansas from August 2013 to June 2014, working on inference for Gaussian systems with D. Nualart and Y. Hu.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific events organisation

8.1.1.1. Member of the organizing committee

- Co-organizer of the Journée Fédération Charles Hermite - entreprises, on 2014/01/23, 150 participants and 44 compagnies coming mainly from Lorraine and few of them from Luxembourg (P. Vallois).

- Organizer of the meeting "Around Hidden Markov Chains", at Inria Grand-Est, on 2014/07/08 (P. Vallois).
- Co-organizer of the workshop "PL^{2}US", Nancy, January 2014 (C. Lacaux).
- Organizer of the weekly seminar of the Probability and Statistics group of the Institut Élie Cartan de Lorraine since October 2014 (R. Azaïs).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

BIGS is a team whose composition included University staff only until October 2014. All members teach numerous courses, ranging from L1 to M2 levels.

- Samy Tindel (192h, Université de Lorraine)
- Thierry Bastogne (192h, Université de Lorraine).
- Sandie Ferrigno (192h, Université de Lorraine)
- Céline Lacaux (192h, Université de Lorraine)
- Jean-Marie Monnez (192h, Université de Lorraine)
- Aurélie Muller-Gueudin (192h, Université de Lorraine)
- Pierre Vallois (192h, University)
- Sophie Wantz-Mézières (192h, IUT)

8.2.2. Supervision

- PhD : Geoffrey Nichil, Provisionnement en assurance non vie pour des contrats à maturité longue et à prime unique - Application à la réforme Solvabilité 2, Université de Lorraine, 2014/12/19. Advisor : P. Vallois, S. Hermmann, M. de Calbiac.

8.2.3. Juries

- PhD comittee : Clémence Chamard-Jovenin, Modélisation du rôle d'ERα36 dans le contrôle de l'initiation et/ou la progression des tumeurs mammaires en réponse aux alkylphénvols, CRAN, Vandœuvre-lès-Nancy, 2014/11/18, examiner: A. Muller-Gueudin.
- PhD Raouf Fakhfakh on 2014/10/02 (Université de Sfax), contribution to the study of Cauchy-Stieltjes kernel families, referee: P. Vallois.
- PhD Raghid Zeineddine, 2014/12/01, Université de Lorraine, Sur des nouvelles formules d'Itô en loi, examiner: P. Vallois.
- PhD : Samuel Ronsin, Régularité et Représentations Localisées de Textures à Phases Aléatoires, MAP 5, Université Paris Descartes, 2014/12/15, examiner: C. Lacaux.

8.3. Popularization

- Fête de la Science, Nancy, École des Mines, 2014/10/17 (A. Muller-Gueudin, S. Tindel, P. Vallois).
- Forum Emploi Maths 4, December 2014 (C. Lacaux).
- Advisor of a group of students, "La main à la Pâte" project, Institut médico-éducatif (IME), Commercy, September-December 2014 (S. Ferrigno).
- Advisor of a group of pupils from Lycée Varoquaux, Tomblaine (Maths en Jeans) (P. Vallois).
- Publication of the short science popularization article "Les huîtres ont des oreilles" in the book "Brèves de maths – Mathématiques de la planète Terre". Editions Nouveau Monde, 2014 (pages 214–215). Authors: Romain Azaïs, Raphaël Coudret and Gilles Durrieu.
- C. Lacaux is member of the *comité national universitaire (section 26)* (2011–2015).

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

[1] 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, n⁰ 4, pp. 950–969 [*DOI : 10.1111/SJOS.12076*], https://hal.archives-ouvertes.fr/hal-01103700

[2] R. AZAÏS, A. GENADOT. *Semi-parametric inference for the absorption features of a growth-fragmentation model*, in "Test (Springer)", October 2014, 20 p. [*DOI : 10.1007/S11749-014-0410-6*], https://hal.inria.fr/hal-01079799

[3] S. COHEN, F. PANLOUP, S. TINDEL. *Approximation of stationary solutions to SDEs driven by multiplicative fractional noise*, in "Stochastic Processes and their Applications", March 2014, vol. 124, n⁰ 3, pp. 1197-1225 [*DOI : 10.1016/J.SPA.2013.11.004*], https://hal.archives-ouvertes.fr/hal-00754368

[4] C. LACAUX, A. MULLER, R. RANTA, S. TINDEL. *Convergence and performance of the peeling wavelet denoising algorithm*, in "Metrika", May 2014, vol. 77, n⁰ 4, pp. 509-537 [*DOI : 10.1007/S00184-013-0451-Y*], https://hal.archives-ouvertes.fr/hal-00903593

[5] B. LALLOUÉ, J.-M. MONNEZ, C. PADILLA, W. KIHAL, D. ZMIROU-NAVIER, S. DEGUEN. *Data analysis techniques: a tool for cumulative exposure assessment*, in "Journal of Exposure Science and Environmental Epidemiology", 2015, vol. 25, pp. 222–230, epub ahead of print [*DOI : 10.1038/JES.2014.66*], https://hal.archives-ouvertes.fr/hal-01069587

[6] M. PERNOT, N. P. E. BARRY, T. BASTOGNE, C. FROCHOT, M. BARBERI-HEYOB, B. THERRIEN. *Rational design of an arene ruthenium chlorin conjugate for in vivo anticancer activity*, in "Inorganica Chimica Acta", April 2014, vol. 414, pp. 134-140 [*DOI : 10.1016/J.ICA.2014.01.048*], https://hal.archives-ouvertes.fr/hal-00967061

Invited Conferences

[7] M. BARBERI-HEYOB, T. BASTOGNE, M. THOMASSIN, R. VANDERESSE, O. TILLEMENT, F. LUX, C. FROCHOT. *Nanoparticles and radiation therapies: Main issues*, in "1er Congrès Annuel SFNano", Nancy, France, December 2014, https://hal.archives-ouvertes.fr/hal-01086007

International Conferences with Proceedings

[8] T. BASTOGNE, J. BRUNIAUX, F. DE CRÉCY, F. NAVARRO. *Empirical model-based identification of critical quality attributes in the preclinical design of nanostructured lipid carriers*, in "7th European Summit for Clinical Nanomedicine and Targeted Medicine, CLINAM 7", Basel, Switzerland, European Foundation for Clinical Nanomedicine, June 2014, https://hal.archives-ouvertes.fr/hal-01078944

[9] J. BRUNIAUX, E. SULPICE, F. MITTLER, M. MENNETEAU, T. BASTOGNE, I. TEXIER, X. GIDROL, F. NAVARRO. *Novel nanostructured lipid carriers dedicated to nucleic acid delivery for RNAi purposes*, in "41st Annual Meeting & Exposition of the Controlled Release Society", Chicago, Illinois, United States, July 2014, https://hal.archives-ouvertes.fr/hal-01078930

[10] J. Bruniaux, E. Sulpice, F. Mittler, M. Menneteau, T. Bastogne, I. Texier, X. Gidrol, F. Navarro. *Novel nanostructured lipid carriers dedicated to nucleic acid delivery for RNAi purposes*, in "Nanotech Advanced Materials and Applications", Washington DC, United States, June 2014, https://hal.archives-ouvertes.fr/hal-01078933

[11] M. Toussaint, F. Auger, N. Durieux, M. Thomassin, T. Bastogne, C. Frochot, S. Mordon, M. Barberi-Heyob. *Magnetic resonance imaging monitoring of interstitial photodynamic therapy for high grade glioma*, in "10th International Symposium on Photodynamic Therapy and Photodiagnosis in Clinical Practice", Brixen, Italy, October 2014, Présentation Poster, https://hal.archives-ouvertes.fr/hal-01079788

[12] M. Toussaint, M. Thomassin, M. Amouroux, T. Bastogne, C. Frochot, S. Mordon, M. Barberi-Heyob. *Optimization of illumination in interstitial photodynamic therapy for high-grade brain tumors*, in "ICPA - International Congress on Photodynamic Applications", Dundee, United Kingdom, May 2014, https://hal.archives-ouvertes.fr/hal-01079400

[13] J.-B. Tylcz, T. Bastogne, A. Bourguignon, S. Pinel, M. Toussaint, C. Frochot, M. Barberi-Heyob. *Real-time control of photobleaching trajectory during photodynamic therapy*, in "ICPA - International Congress on Photodynamic Applications", Dundee, United Kingdom, May 2014, https://hal.archives-ouvertes.fr/hal-01078940

Scientific Books (or Scientific Book chapters)

[14] T. Bastogne, P. Vallois. *An extension of the target theory in biology applied to system reliability*, in "Models and Methods in Economics and Management Science: Essays in Honor of Charles S. Tapiero", F. E. Ouardighi, K. Kogan (editors), International Series in Operations Research & Management Science, Springer, 2014, pp. 155-181, https://hal.archives-ouvertes.fr/hal-00911872

Research Reports

[15] F. Baudoin, E. Nualart, C. Ouyang, S. Tindel. *On probability laws of solutions to differential systems driven by a fractional Brownian motion*, January 2014, https://hal.archives-ouvertes.fr/hal-00931118

[16] M. Gubinelli, S. Tindel, I. Torrecilla. *Controlled viscosity solutions of fully nonlinear rough PDEs*, March 2014, https://hal.archives-ouvertes.fr/hal-00958201

[17] Y. Hu, J. Huang, D. Nualart, S. Tindel. *Stochastic heat equations with general multiplicative Gaussian noises: Hölder continuity and intermittency*, February 2014, https://hal.archives-ouvertes.fr/hal-00944904

[18] Y. Hu, D. Nualart, S. Tindel, F. Xu. *Density convergence in the Breuer-Major theorem for Gaussian stationary sequences*, March 2014, https://hal.archives-ouvertes.fr/hal-00959089

Other Publications

[19] H. Biermé, C. Lacaux. *Modulus of continuity of some conditionally sub-Gaussian fields, application to stable random fields*, 2014, Accepted in Bernoulli, https://hal.archives-ouvertes.fr/hal-00780684

[20] C. Chabriac, A. Lagnoux, S. Mercier, P. Vallois. *Elements related to the largest complete excursion of a reflected BM stopped at a fixed time. Application to local score*, June 2014, https://hal.archives-ouvertes.fr/hal-00857402

[21] S. FERRIGNO, B. FOLIGUET, M. MAUMY-BERTRAND, A. MULLER-GUEUDIN. *Certainty bands for the conditional cumulative distribution function and applications*, May 2014, 25 pages, https://hal.archives-ouvertes.fr/hal-01025443

[22] J.-S. GIET, P. VALLOIS, S. WANTZ-MÉZÌERES. *The logistic S.D.E*, November 2014, https://hal.archives-ouvertes.fr/hal-01096323

[23] C. LACAUX, G. SAMORODNITSKY. *Time-changed Extremal Process as a Random Sup Measure*, October 2014, https://hal.archives-ouvertes.fr/hal-01102343

[24] M. MAUMY-BERTRAND, A. MULLER-GUEUDIN. *Certainty Bands for the Conditional Cumulative Distribution Function and Applications*, September 2014, https://hal.archives-ouvertes.fr/hal-01096041

[25] I.-C. MORARESCU, S. MARTIN, A. GIRARD, A. MULLER-GUEUDIN. *Coordination in networks of linear impulsive agents*, January 2015, https://hal.archives-ouvertes.fr/hal-01096071

References in notes

[26] 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

[27] 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$^{\text{o}}$ 2, pp. 149–157, http://dx.doi.org/10.3166/ejc.14.149-157

[28] A. BENASSI, S. JAFFARD, D. ROUX. *Elliptic Gaussian random processes*, in "Rev. Mat. Iberoamericana", 1997, vol. 13, n$^{\text{o}}$ 1, pp. 19–90

[29] J. O. BERGER. *Statistical decision theory and Bayesian analysis*, Springer Series in Statistics, Springer-VerlagNew York, 1993, xvi+617 p. , Corrected reprint of the second (1985) edition

[30] H. BIERMÉ, C. LACAUX. *Hölder regularity for operator scaling stable random fields*, in "Stochastic Process. Appl.", 2009, vol. 119, n$^{\text{o}}$ 7, pp. 2222–2248, http://dx.doi.org/10.1016/j.spa.2008.10.008

[31] H. BIERMÉ, C. LACAUX, H.-P. SCHEFFLER. *Multi-operator Scaling Random Fields*, in "Stochastic Processes and their Applications", 2011, vol. 121, n$^{\text{o}}$ 11, pp. 2642-2677, MAP5 2011-01 [*DOI* : 10.1016/J.SPA.2011.07.002], http://hal.archives-ouvertes.fr/hal-00551707/en/

[32] 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$^{\text{o}}$ 2, pp. 253–273, http://dx.doi.org/10.1112/blms/bdn122

[33] B. BRUNET-IMBAULT, G. LEMINEUR, C. CHAPPARD, R. HARBA, CLAUDE-LAURENT. BENHAMOU. *A new anisotropy index on trabecular bone radiographic images using the fast Fourier transform*, in "BMC Med. Imaging", 2005, vol. 5, n$^{\text{o}}$ 1, 4 p.

[34] M. CARLETTI. *Mean-square stability of a stochastic model for bacteriophage infection with time delays*, in "Math. Biosci.", 2007, vol. 210, n[o] 2, pp. 395–414, http://dx.doi.org/10.1016/j.mbs.2007.05.009

[35] 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

[36] S. COHEN, C. LACAUX, M. LEDOUX. *A general framework for simulation of fractional fields*, in "Stochastic Process. Appl.", 2008, vol. 118, n[o] 9, pp. 1489–1517, http://dx.doi.org/10.1016/j.spa.2007.09.008

[37] A. CRUDU, A. DEBUSSCHE, O. RADULESCU. *Hybrid stochastic simplifications for multiscale gene networks*, in "BMC Systems Biology", 2009, vol. 3, 89 p. , http://hal.inria.fr/inria-00431227/en/

[38] A. DEYA, S. TINDEL. *Rough Volterra equations. I. The algebraic integration setting*, in "Stoch. Dyn.", 2009, vol. 9, n[o] 3, pp. 437–477, http://dx.doi.org/10.1142/S0219493709002737

[39] S. DOBRE, T. BASTOGNE, M. BARBERI-HEYOB, A. RICHARD. *Practical identifiability of photophysical parameters in photodynamic therapy*, in "29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC'2007", Lyon, France, August 2007, http://hal.archives-ouvertes.fr/hal-00167475/en/

[40] G. B. DURHAM, A. R. GALLANT. *Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes*, in "J. Bus. Econom. Statist.", 2002, vol. 20, n° 3, pp. 297–338, With comments and a reply by the authors, http://dx.doi.org/10.1198/073500102288618397

[41] S. FERRIGNO, G. DUCHARME. *Un test d'adéquation global pour la fonction de répartition conditionnelle*, in "C. R. Math. Acad. Sci. Paris", 2005, vol. 341, n° 5, pp. 313–316, http://dx.doi.org/10.1016/j.crma.2005.07.003

[42] S. FERRIGNO, G. DUCHARME. *Un choix de fenêtre optimal en estimation polynomiale locale de la fonction de répartition conditionnelle*, in "C. R. Math. Acad. Sci. Paris", 2008, vol. 346, n° 1-2, pp. 83–86, http://dx.doi.org/10.1016/j.crma.2007.11.026

[43] S. FERRIGNO, M. MAUMY-BERTRAND, A. MULLER-GUEUDIN. *Uniform law of the logarithm for the conditional distribution function and application to certainty bands*, September 2011, http://hal.inria.fr/hal-00641027

[44] M. GUBINELLI, S. TINDEL. *Rough evolution equations*, in "Ann. Probab.", 2010, vol. 38, n° 1, pp. 1–75, http://dx.doi.org/10.1214/08-AOP437

[45] Y. A. KUTOYANTS. *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London Ltd.London, 2004, xiv+481 p.

[46] C. LACAUX. *Real Harmonizable Multifractional Lévy Motions*, in "Ann. Inst. Poincaré.", 2004, vol. 40, n° 3, pp. 259–277

[47] C. LACAUX. *Series representation and simulation of multifractional Lévy motions*, in "Adv. in Appl. Probab.", 2004, vol. 36, n° 1, pp. 171–197, http://dx.doi.org/10.1239/aap/1077134469

[48] 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

[49] 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/

[50] 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

[51] H. LIERO. *Testing homoscedasticity in nonparametric regression*, in "Journal of Nonparametric Statistics", 2003, vol. 15, n° 1, pp. 31-51, http://www.tandfonline.com/doi/abs/10.1080/10485250306038

[52] T. LYONS, Z. QIAN. *System control and rough paths*, Oxford mathematical monographs, Clarendon Press, 2002, http://books.google.com/books?id=H9fRQNIngZYC

[53] J. MACQUEEN. *Some methods for classification and analysis of multivariate observations.*, 1967, Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281-297 (1967).

[54] J.-M. MONNEZ. *Approximation stochastique en analyse factorielle multiple*, in "Ann. I.S.U.P.", 2006, vol. 50, n° 3, pp. 27–45

[55] J.-M. MONNEZ. *Stochastic approximation of the factors of a generalized canonical correlation analysis*, in "Statist. Probab. Lett.", 2008, vol. 78, n° 14, pp. 2210–2216, http://dx.doi.org/10.1016/j.spl.2008.01.088

[56] J.-M. MONNEZ. *Convergence d'un processus d'approximation stochastique en analyse factorielle*, in "Publ. Inst. Statist. Univ. Paris", 1994, vol. 38, n° 1, pp. 37–55

[57] D. MÁRQUEZ-CARRERAS, C. ROVIRA, S. TINDEL. *Asymptotic behavior of the magnetization for the perceptron model*, in "Ann. Inst. H. Poincaré Probab. Statist.", 2006, vol. 42, n° 3, pp. 327–342, http://dx.doi.org/10.1016/j.anihpb.2005.04.005

[58] 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° 1, pp. 157–174, http://dx.doi.org/10.1214/07-AIHP159

[59] 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

[60] S. TINDEL. *Quenched large deviation principle for the overlap of a p-spins system*, in "J. Statist. Phys.", 2003, vol. 110, n° 1-2, pp. 51–72, http://dx.doi.org/10.1023/A:1021062510565

[61] C. A. TUDOR, F. G. VIENS. *Statistical aspects of the fractional stochastic calculus*, in "Ann. Statist.", 2007, vol. 35, n° 3, pp. 1183–1212, http://dx.doi.org/10.1214/009053606000001541

[62] É. WALTER, L. PRONZATO. *Identification of parametric models*, Communications and Control Engineering Series, Springer-VerlagBerlin, 1997, xviii+413 p. , From experimental data, Translated from the 1994 French original and revised by the authors, with the help of John Norton