IN PARTNERSHIP WITH:
CNRS

Université de Lorraine

Activity Report 2012

Project-Team BIGS

Biology, genetics and statistics

IN COLLABORATION WITH: Institut Elie Cartan Nancy (IECN)

RESEARCH CENTER
Nancy - Grand Est

THEME
Observation, Modeling, and Control for Life Sciences

Table of contents

Project-Team BIGS

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

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

1. Members

Faculty Members

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

Administrative Assistant

Sophie Drouot [shared with Score, Tosca, Veridis]

Others

Mireia Besalú [Postdoc Fellowship, Inria-IECN]
Rémi Bonidal [PhD student, LORIA-IECN]
Benoît Lalloué [PhD student, EHESP-IECN]
Jean-Louis Marchand [Postdoc Fellowship, IECN]
Laura Vinckenbosch [Postdoc Fellowship, Inria-IECN]

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 of Nancy (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 Cramer-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), 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 (i) improve the response reproducibility, (ii) help biologists and chemists in the design of new photosensitizing agents and (iii) 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: (i) Bacteriophage systems (ii) Random fluctuation of nanoparticles. (iii) Automatic detection of osteoporosis.

2.2. Highlights of the Year

For 2012 we stress the following noticeable events:

- HdR defense of Céline Lacaux, 12/6 (see [1]).
- Cybernano, an incubating start-up specialized in nano-cancerology created by Thierry Bastogne, has received the "emergence" award in 2012 from the French Research ministry for the creation of start-up based on innovative technology.

3. Scientific Foundations

3.1. Online data analysis

Participants: J-M. Monnez, R. Bar, 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 today, 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 [61], 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 [58]), which allows to approximate eigenvectors in a stepwise manner. A systematic study of Principal Component and Factorial Analysis has then been leaded by Monnez in the series of papers [64], [62], [63], 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. [59]). 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

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

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 [46], [47], we address the following questions:

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

We also obtain sharp estimates on the conditional distribution function in [48].

3.3. Stochastic modeling for complex and biological systems

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 [31] as well as mixed effects for parameter estimation.

• Another important aspect to quantify for PDT 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 leaded us to introduce a new model based on multi-state Markov chains arguments [10], 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 [39]. 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 [65], [68]) 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 [42], 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 [60], but also invoking the Russo-Vallois integration techniques [67]. The specific issue of Volterra equations driven by fBm, which is central for the subdiffusion within proteins problem, is addressed in [43].

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 [38]). Moreover, the fractional Brownian motion cannot be used to model some phenomena for which the regularity varies with time. Hence, many generalization (gaussian or not) of this model has been recently proposed, see for instance [32] for some Gaussian locally self-similar fields, [54] for some non-Gaussian models, [36] for anisotropic models.

Our team has thus contributed [41], [55], [54], [56], [66] and still contributes [35], [37], [36], [57], [49] 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

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 [72] 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 [33] 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. (i) 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 [31], 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.monolix.org) developed 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. (ii) 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 [44]. 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. (iii) 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 [53] 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 [45].

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 [30], [40]. 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 [69], dealing with some very particular families of equations, which do not cover the cases of interest for us.

4. Application Domains

4.1. Data analysis and local regression

Our expertise in data analysis and advanced statistics methods has given raise no a wide number of interdisciplinary collaborations. Among those, here are the most challenging at a scientific level:

(i) Health inequalities: We have recently developed a statistical procedure in order to create a neighborhood socioeconomic index and investigate its influence on health inequalities. The study setting is composed with 3 major French metropolitan areas (Lille, Lyon and Marseille), and we collaborate for this project with a medical team at EHESP (Ecole des Hautes Etudes en Santé Publique) lead by D. Zmirou (see [19] for further details).

(ii) Fetal pathology: An ongoing work concerning local regression techniques is related to Fetal Biometry, an investigation line suggested by a collaboration between our team and the *Centre de Placentologie et Foetopathologie de la Maternité Régionale de Nancy*, under the direction of Professor Bernard Foliguet. The methods involved in Fetal Biometry are usually based on the comparison of some measured values with the predicted values derived from reference charts or equations in a normal population. However, it happens that maternal and pregnancy characteristics have a significant influence on in-utero Fetal Biometry. We will thus produce some models allowing to construct customized fetal biometric size charts. In order to evaluate them, classical and polynomial regression can be used, but they are not the most appropriate to the kind data we have to handle. Hence, we plan to use local regression estimation in order to perform such an evaluation.

(iii) Cohorts analysis: Some medical teams in Nancy are faced with an overwhelming amount of data, for which a serious statistical assessment is needed. Among those let us mention the INSERM team of Pr. Jean-Louis Guéant and the Inria team Orpailleur (particularly with Marie-Dominique Desvignes and Malika Smail). The goal of this collaboration is to extract biological markers for different diseases (cognitive decline; inflammatory intestinal diseases; liver cancer). To this aim, the INSERM team provides us with several data cohorts with a high number of variables and subjects. As in many instances in Biostatistics, one is then faced with a very high dimensional data, from which we hope to extract a reduced number of significant variables allowing to predict the cardiovascular risk accurately. Moreover, these characters should be meaningful to practitioners. The objective for us is thus to design an appropriate variable selection, plus a classification procedure in this demanding context. Let us highlight an original feature of this collaboration: it combines our own data analysis techniques with those developed by the Orpailleur team, based on symbolic tools. We hope that this experience will enrich both points of view and give raise to new methods of data analysis.

4.2. Estimation for complex and biological systems

Our main application for this line of investigation is the photodynamic therapy developed by T. Bastogne. We shall also focus on bacteriophage therapies and subdiffusion within molecules.

(i) Photodynamic therapy. One of the main application we have in mind for our identification problems is to model photodynamic therapy. This promising cancer treatment involves selective uptake and retention of a photosensitive drug in a tumor, followed by irradiation with light at an appropriate wavelength. Photosensitizers are photoactive compounds such as for instance porphyrins and chlorins. The activated photosensitizer is thought to produce singlet oxygen at high doses and thereby to initiate apoptotic and necrotic death of tumor. Due to the lack of response reproducibility, the complexity of interactions between physical, chemical and biological aspects and the high cost of experiments, there is a real demand in good mathematical and physical models which might help to better control and understand PDT responses. We are particularly concerned with modeling the drug uptake into cancer cells, the photoreactions induced by light exposition and tumor growth kinetics.

(ii) Bacteriophage systems. A collaboration between our team, the Mathematics and the Genetics and Microbiology Departments at the *Universitat Autònoma de Barcelona* (UAB) is being set up, focusing on probabilistic aspects of bacteriophage therapies for animal diseases like hemorrhagic septicemia in cattle or atrophic rhinitis in swine. This kind of therapy consists in inoculating a (benign) virus to animals in order to kill the bacteria known to be responsible of the disease. It was in use in the Soviet Union until the 80s, and is now re-emerging, still at an experimental level, due to the progressive slowdown in antibiotic efficiency.

Within this context, our analysis of a noisy predator-prey competition modeling the treatment helps to calibrate and to understand better the behavior of the system in terms of fluctuations around an equilibrium. Note that our preliminary contacts with the Genetics and Microbiology Departments at UAB also open the way to a particle model in order to represent the couple bacteria/virus living on a surface.

(iii) Subdiffusion into molecules. Our purpose here is a better understanding of the phenomena observed in nanoscale Biophysics, as explained in the series of papers [52]. The technological advances in nanoscale technologies allow the observation of single molecules, and thus the description of newly observed phenomenon. A typical example of this new kind of observation is given by the fluctuations in the folding of a protein-enzyme compound called *Fre*, which is involved in the DNA synthesis of the (canonical) bacterium *E. Coli*.

More specifically, the paper [52] advocates for modeling this folding fluctuations by means of a Volterra type equation driven by a fractional Brownian motion. This convincing model is based on some experimental and physical evidences, and have also been observed in a wide number of recent biological experiments. However, the model exhibited in [52] also raises some unsolved questions: some stochastic equations appearing in the models are not properly defined and their long time behavior is still mysterious. The lack of a method in order to simulate and estimate coefficients of these equations on a solid mathematical ground should also be mentioned. This is the kind of topic we wish to address, for which a preliminary contact with S. Kou and N. Pillai (Princeton University, USA) has been established.

(iv) Osteoporosis. During the year 2011-2012, C. Lacaux has been visiting the MAP 5 (Paris Descartes University) laboratory and joined the ANR Project MATAIM (Modèles Anisotropes de Textures. Applications à l'Imagerie Médicale). This project, which involves both mathematicians and practitioners, is in particular interested in the osteoporosis diagnostic. The paper [34] is a first step in the direction of modeling trabecular bone x-ray images by some operator scaling fields. Actually the estimation of the matrix, which characterizes the anisotropy of the model, is crucial for practical purposes. Hermine Biermé (Paris Descartes Univesity) and Céline Lacaux are working on this problem using quadratic variations. Once the problem of estimation is solved, they plan a comparison of the theoretical model with real data provided by our Biologist colleagues of the MATAIM project. If the model corresponds to real data (as suggested in [34]), this approach may help for the diagnostic of osteoporosis: a numerical study has to be performed in order to find the parameter value which characterizes osteoporosis.

5. Software

5.1. Light diffusion into tissues

We are currently considering the possibility to implement our Matlab algorithms concerning light diffusion into tissues into the Matlab toolbox *Contsid*, developed by the System Identification team of the CRAN (http://www.iris.cran.uhp-nancy.fr/contsid/).

5.2. Online data analysis

A R package performing most of the methods of factorial analysis in an online way is under development 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 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.

5.3. Socio-economic index

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.

6. New Results

6.1. Modern methods of data analysis

Participants: R. Bar, B. Lalloué, J-M. Monnez, C. Padilla, D. Zmirou, S. Deguen.

In 2012, our contributions to data analysis in a Biological context are twofold:

- At a theoretical level, we have kept on working on the so-called online data analysis alluded to at the *Scientific Foundations* Section. Specifically we have carried on in [15] (see also [4]) the analysis of data whose characteristics such as mathematical expectation or covariance matrix may vary with time, a problem which arises very naturally in this context. Moreover, in order to save computation time and thus take into account more data, a method considering several data at each step (we talk about data blocks) is proposed. This technique can also be useful if data are sent and received block-wise. In parallel, a R package performing most of the methods of factorial analysis in an online way is under development.
- At a practical level, our efforts have focused (cf. [19]) on an interesting study concerning the construction of a socio-economic neighborhood index which might quantify health inequalities. While several socio-economic indices already exist in this application field, most of them are very simple both in term of methodological construction and of number of variables taken into account, and only a few use data mining techniques. In order to exploit the large data sets of socio-economic variables provided by censuses and create neighborhood socio-economic indices yielding a better highlight of social health inequalities, a procedure was set in order to automatically select the best indicators in a set of socio-economic variables and synthesize them in a quantitative index. Application to three French metropolitan areas allowed testing the procedure and confirming both its reproducibility on various urban areas and the quality of the neighborhood socio-economic indices we had created (according to field experts and study partners). In this context, our expertise in data analysis allows for a good prediction by means of rigorous methods. Eventually, in order to simplify the application of the creation procedure of a socio-economic index for non-statisticians, a R package called SesIndexCreatoR was created to implement it.
- Publication of the sharp results obtained in [8] on local regression techniques.

6.2. Tumor growth modeling

Participants: R. Keinj, T. Bastogne, P. Vallois.

Up to now, the treatment planning systems used in radiotherapy only use mathematical models to describe the delivery of physical doses of radiation within biological tissues but cannot accurately predict the biological damages caused by such treatments. One important bottleneck is to account for the cell damage heterogeneity in the treated tumor. To this aim we firstly introduced in [51] a stochastic model based on multi-state Markov chains able to describe both treatment damage and cell reparation process.

More recently, we have proposed another model describing the lifespan of heterogenous tumors treated by radiotherapy. It is a bi-scale model in which the cell and tumor lifespans are represented by random variables. First and second-order moments, as well as the cumulative distribution functions and confidence intervals are expressed for the two lifespans with respect to the model parameters. One interesting result is that the mean value of the tumor lifespan can be approached by a logarithmic function of the initial cancer cell number. Moreover, we show that TCP (Tumor Control Probability) and NTCP (Normal Tissue Complication Probability), used in radiotherapy to evaluate, optimize and compare treatment plans, can be derived from the tumor lifespan and the surrounding healthy tissue respectively. Finally, we propose a ROC curve, entitled ECT (Efficiency-Complication Trade-off), suited to the selection by clinicians of the appropriate treatment planning (see [10]).

One difference between photodynamic therapy (PDT) and radiotherapy (RT) is the irradiation signal (X ray in RT and light beam in PDT). Another one is the treatment planning: 10 to 30 daily sessions of treatment in RT against only one for PDT. To adapt the previous model to PDT, a continuous-time version was developed and proposed in [18]. The model has been implemented into Matlab and numerical simulations have emphasized the effects of the model parameters on the model output.

In the framework of a new collaboration with S. Niclou (NorLux Neuro-Oncology Laboratory, Department of Oncology, Centre de Recherche Public de la Santé, Luxembourg), we have extended our stochastic model of cell damage to describe the phenotypic heterogeneity in brain tumors. Preliminary results have recently been presented in [16]. Cancer stem cell (CSC) hypothesis suggests that tumor progression and recurrence rely on a small subpopulation of cancer cells with stem-like properties. The unresolved question is whether cancer stem cells lead to organisation of intratumoral phenotypic heterogeneity by hierarchical differentiation events or whether they represent one of the transitory phenotypic states. This is crucial not only for our understanding of tumor progression, but also for the successful design of novel therapeutic strategies targeting CSCs. Let us also highlight the fact that those studies are related to a more application oriented research synthesized in [3], [13], [21]

6.3. Piecewise deterministic Markov processes

Participants: A. Crudu, A. Debussche, A. Muller-Gueudin, O. Radulescu.

Piecewise deterministic Markov processes are models which feature in a prominent way in Biomedical applications. They appear in two contributions of our team this year.

(1) *Convergence of stochastic gene networks.* In [24], [5], we propose simplified models for the stochastic dynamics of gene network models arising in molecular biology. Those gene networks are classically modeled by Markov jump processes, which are extremely time consuming. To overcome this drawback, we study the asymptotic behavior of multiscale stochastic gene networks using weak limits of Markov jump processes.

We consider a set of chemical reactions R_r, $r \in \mathcal{R}$; $\mathcal{R}$ is supposed to be finite. These reactions involve species indexed by a set $S = 1, \cdots, M$, the number of molecules of the species i is denoted by n_i and $X \in \mathbb{N}^M$ is the vector consisting of the n_i's. Each reaction R_r has a rate $\lambda_r(X)$ which depends on the state of the system, described by X and corresponds to a change $X \to X + \gamma_r$, $\gamma_r \in \mathbb{Z}^M$.

Mathematically, this evolution can be described by the following Markov jump process. It is based on a sequence $(\tau_k)_{k\geq 1}$ of random waiting times with exponential distribution. Setting $T_0 = 0$, $T_i = \tau_1 + \cdots + \tau_i$, X is constant on $[T_{i-1}, T_i)$ and has a jump at T_i. The parameter of τ_i is given by $\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$:

$$\mathbf{P}(\tau_i > t) = \exp\left(-\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))t\right).$$

At time T_i, a reaction $r \in \mathcal{R}$ is chosen with probability $\lambda_r(X(T_{i-1}))/\sum_{r\in\mathcal{R}} \lambda_r(X(T_{i-1}))$ and the state changes according to $X \to X + \gamma_r$: $X(T_i) = X(T_{i-1}) + \gamma_r$. This Markov process has the following generator:

$$Af(X) = \sum_{r\in\mathcal{R}} [f(X + \gamma_r) - f(X)]\, \lambda_r(X).$$

In the applications we have in mind, the numbers of molecules have different scales. Some of the molecules are in small numbers and some are in large numbers. Accordingly, we split the set of species into two sets C and D with cardinals M_C and M_D. This induces the decomposition $X = (X_C, X_D)$, $\gamma_r = (\gamma_r^C, \gamma_r^D)$. For $i \in D$, n_i is of order 1 while for $i \in C$, n_i is proportional to N where N is a large number. For $i \in C$, setting $\widetilde{n}_i = n_i/N$, $\widetilde{n}_i$ is of order 1. We define $x_C = X_C/N$ and $x = (x_C, X_D)$.

For this kind of system, we are able to give in [5] some relevant information on the asymptotic regime $N \to \infty$ when different type of reactions are involved. Depending on the time and concentration scales of the system we distinguish four types of limits:

- Continuous piecewise deterministic processes (PDP) with switching.
- PDP with jumps in the continuous variables.
- Averaged PDP.
- PDP with singular switching.

We justify rigorously the convergence for the four types of limits.

(2) *Variable length Markov chains.* A classical random walk $(S_n,\, n \in \mathbb{N})$ is defined by $S_n := \sum_{k=0}^{n} X_k$, where (X_k) are i.i.d. When the increments $(X_k)_{k\in\mathbb{N}}$ are a one-order Markov chain, a short memory is introduced in the dynamics of (S_n). This so-called "persistent" random walk is no longer Markovian and, under suitable conditions, the rescaled process converges towards the integrated telegraph noise (ITN) as the time-scale and space-scale parameters tend to zero (see [70], [71], [50]). The ITN process is effectively non-Markovian too. In [28] our aim has been to consider persistent random walks (S_t) whose increments are Markov chains with variable order which can be infinite.

Associated with a process (X_n) which takes its values in a finite set, we consider an integer valued process (M_n) so that (X_n, M_n) is Markov and M_n measures the size of the memory at time n. This variable memory is justified by a one-to-one correspondence between (X_n) and a suitable Variable Length Markov Chain (VLMC), since for a VLMC the dependency from the past can be unbounded. We prove in [28] that, under a suitable rescaling, (S_n, X_n, M_n) converges in distribution towards a time continuous process $(S^0(t), X(t), M(t))$. The process $(S^0(t))$ is a semi-Markov and Piecewise Deterministic Markov Process whose paths are piecewise linear.

Observe that, though our study in [28] is made at a theoretical level, it leads to potentially interesting applications in growth models for tumors. This kind of link will be developed in the next future.

6.4. Inference for Gaussian systems

Participants: T. Cass, S. Cohen, M. Hairer, C. Litterer, F. Panloup, L. Quer, S. Tindel.

As mentioned at 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 (see [69]) 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:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\, du + \int_0^t \sigma(\theta; X_u)\, dB_t, \tag{1}$$

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 2012:

- An implementable numerical scheme for equations driven by irregular processes, which is one of the ingredients one needs in order to perform an accurate statistical estimation procedure (see [6]).
- A better understanding of the law of the solution X_t^θ to equation (1), carried out in [25]. This step allows to obtain smoothness of density for our equation of interest in a wide range of contexts, which is an essential prerequisite for a good estimation procedure.
- Another important preliminary step for likelihood estimates for stochastic equations is a good knowledge of their invariant measure in the ergodic case. This is the object of our article [27].
- Finally we have also progressed in our knowledge of noisy differential systems by extending the range of applications of rough paths methods [11], [14].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Start-up project by T. Bastogne:

Industrial partner: CyberBio (Biocybernetics for Cancerology & Nanomedicine).

Status: in incubation.

Comments: Cybernano is an incubating start-up specialized in nano-cancerology, which has received the "emergence" award in 2012 from the French Research ministry for the creation of start-up based on innovative technology. Cybernano proposes innovating products to reduce the cost and control the risk during the preclinical development of nanoparticles in oncology applications. The engineering approach used by this spin-off is strongly based on the use of suited mathematical models.

7.2. Bilateral Grants with Industry

CIFRE PhD grant supervised by P. Vallois:

Industrial partner: Caisse Mutuelle du Crédit Agricole.

Title: Claim reserving for insurance.

PhD thesis of M. Geoffray Nichil.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Co-direction of a PhD thesis by J-M. Monnez:

Partner: Ecole de Hautes Etudes en Santé Publique (Rennes).

Title: Influence of socio-economic and environmental characteristics on infant mortality.

PhD thesis of M. Lalloué.

Regional project leaded by T. Bastogne:

Partners: Contrat de Projets Etat-Région, MISN (Modélisation, Information et Système Numérique), Thème AOC (Analyse, Optimisation et Contrôle).

Title: EMC2 (Experimental design, Modeling and Control in Cancerology).

8.2. National Initiatives

• C. Lacaux is member of the MATAIM (Modèles Anisotropes de Textures. Applications à l'Imagerie Médicale) ANR project, leaded by F. Richard (University of Provence).

• P. Vallois is member of the MASTERIE (Malliavin Stein Random Irregular Equation) ANR project, leaded by F. Russo (ENSTA, Paris).

• T. Bastogne is leader of the MOCOBIO (MOdeling and COntrol of heterogeneous systems in BIOlogy) CNRS-PEPS project.

• T. Bastogne is member of the PDTX (Active Nanoplatforms for Photodynamic Therapy) ANR project, lead by M. Verelst (Université Paul Sabatier, Toulouse)

• T. Bastogne is member of the Nano-VTP (Nanoparticles for Imaging and Vascular Photodynamic Treatment of Brain Tumors) ANR project, lead by M. Barberi-Heyob (Centre de Recherche en Automatique de Nancy, Centre Alexis Vautrin).

• T. Bastogne, C. Lacaux and S. Tindel are members of the OPTIQUE CNRS-PEPS project, lead by M. Thomassin (CRAN) and managed within Inria's framework by BIGS.

8.3. European Initiatives

8.3.1. Collaborations with Major European Organizations

Collaboration 1: Smoothness of density for noisy differential systems

Partner 1: Imperial College, London (UK)

Partner 2: Warwick University (UK)

Subject: Smoothness of density for noisy differential systems

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Visit of D. Nualart (Kansas University) for 1 month, May 2012.

8.4.1.1. Internships

- Raouf Souabni: Simulation of the light propagation in biological tissues. Application to interstitial photodynamic therapy. Advisor: T. Bastogne.
- Yosra Chemli: Applicability of an Exponential-Linear (E-L) model to describe the in vitro cell responses in photodynamic therapy. Advisor: T. Bastogne.
- Kevin Ziegelmeyer: Data Analysis for for liver cirrhosis prediction. Advisor: A. Muller-Gueudin.

9. Dissemination

9.1. Scientific Animation

Conferences organized by our team in 2012:

- STochastic ANalysis days, 9-11 May: A 3 days international meeting gathering some of the best specialists in stochastic analysis and applications (including statistics and fractional fields). Organizers: C. Lacaux, I. Nourdin, S. Tindel.
- Journée Fédération Charles Hermite, 15 October: Modeling for Cancer Therapies. Organizers: W. Blondel, S. Tindel.
- Journée Modélisation des Biomolécules et leurs Interactions, 25-26 October. Organizers: M-D. Devignes, Aurélie Muller-Gueudin.
- Weekly Biostats Seminar at IECN, organized by Aurélie Muller-Gueudin. See http://www.iecn.u-nancy.fr/~muller/gt.html.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

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

9.2.2. Supervision

HDR : C. Lacaux, *Contributions à la notion d'autosimilarité et à l'étude des trajectoires de champs aléatoires*, Universté de Lorraine, 6/12/2012. See document [1].

PhD in progress: R. Bar, *Analyse de donnés en ligne*, from 01/09/2010. Advisor: J-M. Monnez.

PhD in progress: B. Lalloué: *Analyse des données dans l'étude de l'influence de caractéristiques socio-spatiales sur des événements de santé*, from 01/09/2010. Advisor: J-M. Monnez.

PhD in progress: R. Bonidal: *Analyse des systèmes discriminants multi-classes à grande marge*, from 01/09/2009. Advisors: Y. Guermeur, S. Tindel.

PhD in progress: G. Nichil, *Claim reserving for insurance*, from 01/09/2010. Advisors: S. Herrmann (University of Dijon), P. Vallois.

9.2.3. Juries

PhD : S. Bounebache, *Stochastic partial differential equations of parabolic type with singular potential*, Universté de Paris 6, June 2012. Advisor: L. Zambotti. Referee: S. Tindel.

PhD: J. Valentin, *Weak Itô formulae*, Telecom Paristech, June 2012. Advisor: A.S. Üstünel. Referee: S. Tindel.

HDR: B. Saussereau: *Equations aux dérivés partielles stochastiques; Equations différentielles stochastiques dirigées par un mouvement brownien fractionnaire*, Université de Franche Comté, November 2012. Referee: S. Tindel.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] C. LACAUX. *Contribution à la notion d'autosimilarité et à l'étude des trajectoires de champs aléatoires.*, Université de Lorraine, December 2012, Habilitation à Diriger des Recherches, http://hal.inria.fr/tel-00764398.

Articles in International Peer-Reviewed Journals

[2] H. BENACHOUR, T. BASTOGNE, M. TOUSSAINT, Y. CHEMLI, A. SÈVE, C. FROCHOT, F. LUX, O. TILLEMENT, R. VANDERESSE, M. BARBERI-HEYOB. *Real-time monitoring of photocytotoxicity in nanoparticles-based photodynamic therapy: a model-based approach*, in "PLoS ONE", November 2012, vol. 7, n^{o} 11, e48617 [*DOI :* 10.1371/JOURNAL.PONE.0048617], http://hal.inria.fr/hal-00742724.

[3] H. BENACHOUR, A. SÈVE, T. BASTOGNE, C. FROCHOT, R. VANDERESSE, J. JASNIEWSKI, I. MILADI, C. BILLOTEY, O. TILLEMENT, F. LUX, M. BARBERI-HEYOB. *Multifunctional peptide-conjugated hybrid silica nanoparticles for photodynamic therapy and MRI*, in "Theranostics", 2012, vol. 2, n^{o} 9, p. 889-904 [*DOI :* 10.7150/THNO.4754], http://hal.inria.fr/hal-00723908.

[4] H. CARDOT, P. CÉNAC, J.-M. MONNEZ. *A fast and recursive algorithm for clustering large datasets with k-medians*, in "Computational Statistics and Data Analysis", 2012, vol. 56, p. 1434-1449 [*DOI :* 10.1016/J.CSDA.2011.11.019], http://hal.inria.fr/hal-00644683.

[5] A. CRUDU, A. DEBUSSCHE, A. MULLER, O. RADULESCU. *Convergence of stochastic gene networks to hybrid piecewise deterministic processes*, in "Annals of Applied Probability", 2012, vol. 22, n^{o} 5, p. 1822-1859 [*DOI :* 10.1214/11-AAP814], http://hal.inria.fr/hal-00553482.

[6] A. DEYA, A. NEUENKIRCH, S. TINDEL. *A Milstein-type scheme without Levy area terms for SDEs driven by fractional Brownian motion*, in "Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques", 2012, vol. 48, n^{o} 2, p. 518-550, http://hal.inria.fr/hal-00448685.

[7] S. DOBRE, T. BASTOGNE, C. PROFETA, M. BARBERI-HEYOB, A. RICHARD. *Limits of variance-based sensitivity analysis for non- identifiability testing in high dimensional dynamic models*, in "Automatica", August 2012, vol. 48, n^{o} 1, p. 2740-2749 [*DOI :* 10.1016/J.AUTOMATICA.2012.05.004], http://hal.inria.fr/hal-00730316.

[8] G. DUCHARME, S. FERRIGNO. *An Omnibus Test of Goodness-of-fit for Conditional Distribution Functions with Applications to Regression Models*, in "Journal of Statistical Planning and Inference", October 2012, vol. 142, n^{o} 10, p. 2748-2761, http://hal.inria.fr/hal-00641034.

[9] J. ISTAS, C. LACAUX. *On locally self-similar fractional random fields indexed by a manifold*, in "Stochastics An International Journal of Probability and Stochastic Processes: formerly Stochastics and Stochastics Reports", 2013, to be published [*DOI :* 10.1080/17442508.2012.654609], http://hal.inria.fr/hal-00193203.

[10] R. KEINJ, T. BASTOGNE, P. VALLOIS. *Tumor growth modeling based on cell and tumor lifespans*, in "Journal of Theoretical Biology", July 2012, vol. 312, p. 76-86 [*DOI :* 10.1016/J.JTBI.2012.07.008], http://hal.inria.fr/hal-00730315.

[11] J. LEON, S. TINDEL. *Malliavin calculus for fractional delay equations*, in "Journal of Theoretical Probability", 2012, vol. 25, n^{o} 3, p. 854-889, http://hal.inria.fr/hal-00440655.

[12] J. MRIOUAH, C. BOURA, M. THOMASSIN, T. BASTOGNE, D. DUMAS, B. FAIVRE, M. BARBERI-HEYOB. *Tumor vascular responses to antivascular and antiangiogenic strategies: looking for suitable models*, in "Trends in Biotechnology", September 2012, vol. 30, n^{o} 12, p. 649-658 [*DOI :* 10.1016/J.TIBTECH.2012.08.006], http://hal.inria.fr/hal-00726417.

[13] M. PERNOT, T. BASTOGNE, N. P. E. BARRY, B. THERRIEN, G. KOELLENSPERGER, S. HANN, V. RESHETOV, M. BARBERI-HEYOB. *System biology approach for in vivo photodynamic therapy optimization of ruthenium-porphyrin compounds*, in "Journal of Photochemistry and Photobiology B: Biology", December 2012, vol. 117, p. 80-89 [*DOI :* 10.1016/J.JPHOTOBIOL.2012.08.012], http://hal.inria.fr/hal-00724050.

[14] L. QUER-SARDANYONS, S. TINDEL. *Pathwise definition of second order SDEs*, in "Stochastic Processes and their Applications", 2012, vol. 122, n^{o} 2, p. 466-497, http://hal.inria.fr/hal-00531331.

International Conferences with Proceedings

[15] R. BAR, J.-M. MONNEZ. *Analyse canonique généralisée de données séquentielles d'espérance fixe ou variable dans le temps*, in "SFDS - 44èmes journées de Statistique - 2012", Bruxelles, Belgium, May 2012, http://hal.inria.fr/hal-00734606.

[16] T. BASTOGNE, A. GOLEBIEWSKA, F. AZUAJE, S. NICLOU. *Markovian modelling of phenotypic state transitions in glioblastoma cancer cells.*, in "EMBO Conference Series: From Functional Genomics to Systems Biology", Heidelberg, Germany, November 2012, 122, http://hal.inria.fr/hal-00753925.

[17] S. DOBRE, J.-B. TYLCZ, T. BASTOGNE, M. BARBERI-HEYOB. *Global sensitivity analysis and estimation of photophysical parameters from in vivo data in photodynamic therapy*, in "16th IFAC Symposium on System

Identification, SYSID'2012", Bruxelles, Belgium, M. KINNAERT (editor), IFAC, July 2012, p. 1731-1736 [*DOI :* 10.3182/20120711-3-BE-2027.00268], http://hal.inria.fr/hal-00751746.

[18] R. KEINJ, M. THOMASSIN, T. BASTOGNE, P. VALLOIS. *Model of cell damage in single session radiation treatments based on continuous-time Markov chains*, in "8th IFAC Symposium on Biological and Medical Systems, IFAC BMS 2012", Budapest, Hungary, August 2012, CDROM [*DOI :* 10.3182/20120829-3-HU-2029.00081], http://hal.inria.fr/hal-00762658.

[19] B. LALLOUÉ, J.-M. MONNEZ, C. PADILLA, D. ZMIROU-NAVIER, S. DEGUEN. *Méthodologie de Création d'un Indice de Défaveur Contextuelle - Un Outil Permettant l'Analyse des Inégalités Sociales de Santé*, in "SFDS - 44èmes journées de Statistique - 2012", Bruxelles, Belgium, May 2012, Submission, http://hal.inria.fr/hal-00734769.

[20] M. TOUSSAINT, M. THOMASSIN, M. AMOUROUX, E. GALBRUN, W. BLONDEL, T. BASTOGNE, M. BARBERI-HEYOB. *Optimization of illumination in interstitial photodynamic therapy for high-grade brain tumors*, in "9th International Symposium on Photodynamic Therapy and Photodiagnosis in Clinical Practice", Brixen, Italy, October 2012, CDROM, http://hal.inria.fr/hal-00753936.

[21] J.-B. TYLCZ, D. BECHET, T. BASTOGNE, H. GARNIER, M. BARBERI-HEYOB. *System identification of the intrabrain tumoral uptake of multifunctional nanoparticles*, in "8th IFAC Symposium on Biological and Medical Systems, IFAC BMS 2012", Budapest, Hungary, August 2012, CDROM, http://hal.inria.fr/hal-00753965.

[22] J.-B. TYLCZ, D. BECHET, T. BASTOGNE, H. GARNIER, M. BARBERI-HEYOB. *Pharmacokinetic modeling of nanoparticles-based PS vectors into glioblastoma from MRI imaging in PDT*, in "9th International Symposium on Photodynamic Therapy and Photodiagnosis in Clinical Practice", Brixen, Italy, October 2012, CDROM, http://hal.inria.fr/hal-00753928.

National Conferences with Proceeding

[23] J.-M. MONNEZ, R. BAR. *Analyse canonique généralisée d'un flux de données d'espérance variable dans le temps*, in "XIXèmes Rencontres de la Société Francophone de Classification - SFC 2012", Marseille, France, October 2012, http://hal.inria.fr/hal-00750883.

Conferences without Proceedings

[24] A. MULLER-GUEUDIN. *Modélisation de réseaux de régulation de gènes par processus déterministes par morceaux*, in "44emes Journées de la SFDS", BRUXELLES, Belgium, May 2012, 6 pages, http://hal.inria.fr/hal-00757861.

Research Reports

[25] T. CASS, M. HAIRER, C. LITTERER, S. TINDEL. *Smoothness of the density for solutions to Gaussian Rough Differential Equations*, September 2012, http://hal.inria.fr/hal-00732063.

Other Publications

[26] R. BAR, J.-M. MONNEZ. *Analyse Canonique Généralisée de données séquentielles*, 2012, http://hal.inria.fr/hal-00734566.

[27] S. COHEN, F. PANLOUP, S. TINDEL. *Approximation of stationary solutions to SDEs driven by multiplicative fractional noise*, http://hal.inria.fr/hal-00754368.

[28] P. CÉNAC, B. CHAUVIN, S. HERRMANN, P. VALLOIS. *Persistent random walks, variable length Markov chains and piecewise deterministic Markov processes*, July 2012, Submitted for publication, http://hal.inria.fr/hal-00719450.

[29] B. LALLOUÉ, J.-M. MONNEZ, C. PADILLA, N. LE MEUR, D. ZMIROU-NAVIER, S. DEGUEN. *A statistical procedure to create a neighborhood socioeconomic index for health inequalities analysis*, 2012, http://hal.inria.fr/hal-00759868.

References in notes

[30] 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, p. 557?-577.

[31] 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^{o} 2, p. 149–157, http://dx.doi.org/10.3166/ejc.14.149-157.

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

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

[34] H. BIERMÉ, CLAUDE-LAURENT. BENHAMOU, F. RICHARDE. *Parametric estimation for Gaussian operator scaling random fields and anisotropy analysis of bone radiographs texture*, in "Proc. MICCAI 12th Int. Conference, London", 2009.

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

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

[37] 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^{o} 2, p. 253–273, http://dx.doi.org/10.1112/blms/bdn122.

[38] 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^{o} 1, 4.

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

[40] J. F. COEURJOLLY. *Simulation and identification of the fractional brownian motion: a bibliographical and comparative study.*, in "Journal of Statistical Software", 2000, vol. 5, p. 1–53.

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

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

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

[44] 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, CDROM, http://hal.archives-ouvertes.fr/hal-00167475/en/.

[45] 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^{o} 3, p. 297–338, With comments and a reply by the authors, http://dx.doi.org/10.1198/073500102288618397.

[46] 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^{o} 5, p. 313–316, http://dx.doi.org/10.1016/j.crma.2005.07.003.

[47] 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^{o} 1-2, p. 83–86, http://dx.doi.org/10.1016/j.crma.2007.11.026.

[48] 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.

[49] M. GUBINELLI, S. TINDEL. *Rough evolution equations*, in "Ann. Probab.", 2010, vol. 38, n^{o} 1, p. 1–75, http://dx.doi.org/10.1214/08-AOP437.

[50] S. HERRMANN, P. VALLOIS. *From persistent random walk to the telegraph noise*, in "Stoch. Dyn.", 2010, vol. 10, n^{o} 2, p. 161–196, http://dx.doi.org/10.1142/S0219493710002905.

[51] 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^{o} 1, p. 55-62 [*DOI :* 10.1016/J.JTBI.2011.03.025], http://hal.inria.fr/hal-00588935/en.

[52] S. C. KOU. *Stochastic Networks in Nanoscale Biophysics*, in "Journal of the American Statistical Association", 2008, vol. 103, n^{o} 483, p. 961-975, http://pubs.amstat.org/doi/abs/10.1198/016214507000001021.

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

[54] C. LACAUX. *Real Harmonizable Multifractional Lévy Motions*, in "Ann. Inst. Poincaré.", 2004, vol. 40, n^o 3, p. 259–277.

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

[56] C. LACAUX, J.-M. LOUBES. *Hurst exponent estimation of fractional Lévy motion*, in "ALEA Lat. Am. J. Probab. Math. Stat.", 2007, vol. 3, p. 143–164.

[57] 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/.

[58] 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, Vienna, 1974, p. 202–211.

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

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

[61] 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)..

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

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

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

[65] 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^o 3, p. 327–342, http://dx.doi.org/10.1016/j.anihpb.2005.04.005.

[66] 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^o 1, p. 157–174, http://dx.doi.org/10.1214/07-AIHP159.

[67] F. RUSSO, P. VALLOIS. *Elements of stochastic calculus via regularization*, in "Séminaire de Probabilités XL", Berlin, Lecture Notes in Math., Springer, Berlin, 2007, vol. 1899, p. 147–185, http://dx.doi.org/10.1007/978-3-540-71189-6_7.

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

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

[70] P. VALLOIS, C. S. TAPIERO. *Memory-based persistence in a counting random walk process*, in "Phys. A.", 2007, vol. 386, n^{o} 1, p. 303–307, http://dx.doi.org/10.1016/j.physa.2007.08.027.

[71] P. VALLOIS, C. S. TAPIERO. *A claims persistence process and insurance*, in "Insurance Math. Econom.", 2009, vol. 44, n^{o} 3, p. 367–373, http://dx.doi.org/10.1016/j.insmatheco.2008.11.009.

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