



IN PARTNERSHIP WITH:  
**CNRS**

**Université de Lorraine**

Activity Report 2011

# 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

<b>1. Members</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>1</b>
2.1. Overall Objectives	1
2.2. Highlights	2
<b>3. Scientific Foundations</b>	<b>2</b>
3.1. Online data analysis	2
3.2. Local regression techniques	2
3.3. Stochastic modeling for complex and biological systems	3
3.4. Parameter identifiability and estimation	4
<b>4. Application Domains</b>	<b>5</b>
4.1. Data analysis and local regression	5
4.2. Estimation for complex and biological systems	5
<b>5. Software</b>	<b>7</b>
<b>6. New Results</b>	<b>7</b>
6.1. Modern methods of data analysis	7
6.2. Local linear estimator of the conditional distribution function	8
6.3. Markovian models for tumor growth	8
6.4. A stochastic model for bacteriophage therapies	9
6.5. Convergence of stochastic gene networks	10
6.6. Inference for Gaussian systems	11
6.7. Local self-similarity properties and stable or Gaussian random fields	12
<b>7. Contracts and Grants with Industry</b>	<b>13</b>
7.1. Contracts with Industry	13
7.2. Grants with Industry	13
<b>8. Partnerships and Cooperations</b>	<b>13</b>
8.1. Regional Initiatives	13
8.2. National Initiatives	13
8.3. European Initiatives	14
8.3.1. Collaborations in European Programs, except FP7	14
8.3.2. Major European Organizations with which Bigs has followed Collaborations	14
8.4. International Initiatives	14
8.5. Teaching	14
<b>9. Bibliography</b>	<b>15</b>



# Project-Team BIGS

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

## 1. Members

### Faculty Members

Samy Tindel [Team leader, Professor, Université Nancy 1, HdR]  
Thierry Bastogne [Associate Professor, Université Nancy 1, HdR]  
Sandie Ferrigno [Associate Professor, INPL]  
Céline Lacaux [Associate Professor, INPL]  
Aurélie Muller [Associate Professor, INPL]  
Jean-Marie Monnez [Professor, Université Nancy 2, HdR]  
Pierre Vallois [Professor, Université Nancy 1, HdR]  
Sophie Wantz [Associate Professor, Université Nancy 2]

### Administrative Assistant

Sophie Drouot [shared with Corida, Tosca]

## 2. Overall Objectives

### 2.1. Overall Objectives

BIGS is a team labeled by INRIA, by CNRS and by University Henri Poincaré, via the Institut Élie Cartan of Nancy (UMR 7502 CNRS-INRIA-UHP-INPL-University of Nancy 2). 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

For 2011 we stress the following noticeable events:

- PhD defense of Aurélien Deya (supervisor: Samy Tindel).
- PhD defense of Roukaya Keinj (supervisors: Thierry Bastogne and Pierre Vallois).

## 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 [56], 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 [52]), which allows to approximate eigenvectors in a stepwise manner. A systematic study of Principal Component and Factorial Analysis has then been led by Monnez in the series of papers [59], [57], [58], 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. 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  $\epsilon$ . 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. [54]). Procedures testing the normality of the  $\epsilon_i$  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 [43], [44], 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.

### 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 [26] 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 led us to introduce a new model based on multi-state Markov chains arguments [3], 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 [60], [63]) in order to combine convergence to equilibrium for the deterministic system and deviations of the stochastic system. Notice that A. Muller is also working with A. Debussche and O. Radulescu on a related topic [38], 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 [55], but also invoking the Russo-Vallois integration techniques [62]. The specific issue of Volterra equations driven by fBm, which is central for the subdiffusion within proteins problem, is addressed in [40].

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 medicine (see [33]). 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 [27] for some Gaussian locally self-similar fields, [48] for some non-Gaussian models, [31] for anisotropic models.

Our team has thus contributed [36], [49], [48], [50], [61] and still contributes [30], [32], [31], [51], [45] 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 [65] 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 [28] 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 [26], 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 [41]. 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 [47] 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 [42].



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 [25], [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 [64], 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 rise to a wide number of interdisciplinary collaborations. Among those, here are the most challenging at a scientific level:

(i) *Peanut allergy*: In the recent past, a direct application of factorial analysis techniques has been concerned with a study about allergic patients. This project was focusing on allergies to peanut, and aimed at predicting the level of an allergic crisis according to some biological parameters. In this context, no rigorous discriminant analysis had been performed before, and the article [2] has been considered as an achievement in this direction.

(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 of 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 *Stanislas* cohort handled at the *Centre Alexis Vautrin*, which provides a huge amount of data potentially enabling a sharp identification of the biological characters involved in cardiovascular diseases. 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 also mention the starting collaboration with 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 the Stanislas cohort, the objective for us is to design an appropriate variable selection, plus a classification procedure in this demanding context. This work has the originality to combine our own 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 rise 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 [46]. 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 [46] 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 [46] 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 2010-2011, 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 [29] 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 University) 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 [29]), 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. Identification of biological systems

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

## 6. New Results

### 6.1. Modern methods of data analysis

Participants: H. Cardot, P. Cénac, O. Collignon, J-M. Monnez, P. Vallois.

In 2011, 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 the construction of a fast and recursive algorithm for clustering large data sets with the  $k$ -medians methods.
- At a practical level, our efforts have focused on an interesting study concerning peanuts allergy, for which our expertise in data analysis allows for a good prediction of allergy severity by means of rigorous methods.

Let us now describe more precisely our articles:

(i) *A fast and recursive algorithm for clustering large data sets with  $k$ -medians.* Clustering with fast algorithms large samples of high dimensional data is an important challenge in computational statistics. Borrowing ideas from MacQueen [56], who introduced a sequential version of the  $k$ -means algorithm, a new class of recursive stochastic gradient algorithms designed for the  $k$ -medians loss criterion is proposed in [16], [17]. By their recursive nature, these algorithms are very fast and well adapted to deal with large samples of data that are allowed to arrive sequentially. It is proved that the stochastic gradient algorithm converges almost surely to the set of stationary points of the underlying criterion. A particular attention is paid to the averaged versions, which are known to have better performances, and a data-driven procedure that allows automatic selection of the value of the descent step is proposed. The performance of the averaged sequential estimator is compared on a simulation study, both in terms of computation speed and accuracy of the estimations, with more classical partitioning techniques such as  $k$ -means, trimmed  $k$ -means and PAM (partitioning around medoids). Finally, this new on-line clustering technique is illustrated on determining television audience profiles with a sample of more than 5000 individual television audience measured every minute over a period of 24 hours.

(ii) *Discriminant analyses of peanut allergy severity scores.* Peanut allergy is one of the most prevalent food allergies. The possibility of a lethal accidental exposure and the persistence of the disease make it a public health problem. Evaluating the intensity of symptoms is accomplished with a double blind placebo-controlled food challenge (DBPCFC), which scores the severity of reactions and measures the dose of peanut that elicits the first reaction. Since DBPCFC can result in life-threatening responses, we propose in [2] an alternate procedure with the long-term goal of replacing invasive allergy tests. Discriminant analysis of DBPCFC score, the eliciting dose and the first accidental exposure score were performed in 76 allergic patients using 6 immunoassays and 28 skin prick tests. A multiple factorial analysis was performed to assign equal weights to both groups of variables, and predictive models were built by cross-validation with linear discriminant analysis,  $k$ -nearest neighbors, classification and regression trees, penalized support vector machine, stepwise logistic regression and Adaboost methods. We developed an algorithm for simultaneously clustering eliciting doses and selecting discriminant variables. Our main conclusion is that antibody measurements offer information on the allergy severity, especially those directed against *rAra-h1* and *rAra-h3*. Further independent validation of these results and the use of new predictors will help extend this study to clinical practices.

## 6.2. Local linear estimator of the conditional distribution function

Participants: S. Ferrigno, M. Maumy, A. Muller.

Consider  $(X, Y)$ , a random vector defined in  $\mathbb{R} \times \mathbb{R}$ . Here  $Y$  is the variable of interest and  $X$  the concomitant variable. As usual in the statistics literature, we work under the assumption that a sample  $\{(X_i, Y_i)_{1 \leq i \leq n}\}$  of independent and identically replica of  $(X, Y)$  is available.

In order to explain the relationship between the variable of interest  $Y$  and the factor  $X$ , the standard way is to rely on the regression function  $E(Y|X = x)$ . Because of numerous applications, the problem of estimating the regression function has been the subject of considerable interest during the last decades. However, it can be easily argued that the function  $x \mapsto E(Y|X = x)$  alone does not capture the complexity of the relations between  $X$  and  $Y$ .

In order to go one step further in this direction, we have chosen to work with another function. Namely, we study the conditional distribution function  $F(y|X = x) = P(Y \leq y|X = x)$  and a nonparametric estimator associated to this quantity. The distribution function has the advantage of completely characterizing the law of the random variable at stake, allowing to obtain the regression function, the density function, the moments and the quantile function. It should also be noticed that conditional distribution functions are used for the estimation of references curves in medical applications.

At a more technical level, our study is based on a local linear nonparametric estimator of the conditional distribution function instead of the widely spread Nadaraya-Watson estimator. Indeed, it is a well-known fact that the asymptotic bias of the Nadaraya-Watson estimator behaves somehow badly. Observe however that local polynomial techniques are good alternatives. Based on these techniques, here are the steps we have focused on in 2010-2011 :

- Our main result is the uniform law of the logarithm concerning the local linear estimator of the conditional distribution function (see [21]). We investigate convergence in probability and almost sure convergence results.
- The uniform law of the logarithm has then been used to construct uniform asymptotic certainty bands for the conditional distribution function.
- The certainty bands alluded to above have been applied to simulated data.
- A variant of the test has been introduced in [20].

Let us also mention that applications of these theoretical results to survival analysis are currently the object of active research.

## 6.3. Markovian models for tumor growth

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

Our research in this direction includes two contributions in 2011:

- A multinomial model for cell growth allowing to calibrate radiotherapies given in [3].
- A study of tumor growth based on the lifespan of each cell (see [13]).

More specifically, our two contributions can be summarized as follows:

(i) Hit and target models of tumor growth typically assume that all surviving cells have a constant and homogeneous sensitivity during the radiotherapy period. In [3], we propose a multinomial model based on a discrete-time Markov chain, able to take into account cell repair, cell damage heterogeneity and cell proliferation. The proposed model relies on the 'Hit paradigm' and 'Target' theory in radiobiology and assumes that a cancer cell contains  $m$  targets which must be all deactivated to produce cell death. The surviving cell population is then split up into  $m$  categories to introduce the variation of cancer cell radio-sensitivity according to their damage states. Two other parameters have been introduced : the probability  $q$  for a target to be deactivated by radiation and the probability  $r$  for an inactive target in an alive cell to be reactivated.

The parameter  $q$  is related to the radiation dose  $u_0$  through the intrinsic sensitivity of a target to radiation. Moreover, the multinomial model is a generalization of typical hit models. Based on the multinomial model, new expressions of the TCP (Tumor Control Probability) and NTCP (Normal Tissue Complication Probability) have been proposed for nonuniform radiations which permits to deduce the optimal total dose to be delivered. We point out the important influence of the repair parameter  $r$  which could lead to reduce both the total radiation dose to be delivered and the risk of side effects.

(ii) We have proposed in [13] an original approach that expresses the probability distribution of the cancer and normal cells lifespans in terms of the number of dose fractions in radiotherapy. Conversely to previous models that examines the number of surviving cells in the treated population at fixed time instants, our modeling approach better reveals the dynamics of the tumor response.

We start by considering the lifespan of a single cancer cell that behaves as described in [3]. We study this random time by calculating its mean, variance and cumulative distribution function. We then assume that a tumor is a group of independent cells. This allows to define the lifespan of the tumor as the maximum of individual lifespans. When the initial number  $n_0$  of cancer cells is not too large, then we can explicitly calculate the mean, variance and the cumulative distribution function of the tumor lifespan. When  $n_0$  is large, the previous parameters are no longer calculable. However, we are able to show that, under some assumptions, the mean lifespan of the tumor behaves as a logarithmic function of the initial number  $n_0$ . The second goal is to show that TCP and NTCP can be completely formulated with respect to the tumor and normal tissue lifespans. These expressions of TCP and NTCP are finally used to propose a ROC curve, called ECT (Efficiency-Complication Trade-off), suited to the determination of the appropriate treatment schedule. This synthetic representation summarizes both efficiency and complication of the treatment. Moreover, it allows several possibilities of choice for the radiotherapist : treatment efficiency, priority to safety of normal tissue, or a trade-off between them.

## 6.4. A stochastic model for bacteriophage therapies

Participants: X. Bardina, D. Bascompte, C. Rovira, S. Tindel.

In the last years Bacteriophage therapies are attracting the attention of several scientific studies. They can be a new and powerful tool to treat bacterial infections or to prevent them applying the treatment to animals such as poultry or swine. Very roughly speaking, they consist in inoculating a (benign) virus in order to kill the bacteria known to be responsible of a certain disease. This kind of treatment is known since the beginning of the 20th century, but has been in disuse in the Western world, erased by antibiotic therapies. However, a small activity in this domain has survived in the USSR, and it is now re-emerging (at least at an experimental level). Among the reasons of this re-emersion we can find the progressive slowdown in antibiotic efficiency (antibiotic resistance). Reported recent experiments include animal diseases like hemorrhagic septicemia in cattle or atrophic rhinitis in swine, and a need for suitable mathematical models is now expressed by the community.

Let us be a little more specific about the (lytic) bacteriophage mechanism: after attachment, the virus' genetic material penetrates into the bacteria and use the host's replication mechanism to self-replicate. Once this is done, the bacteria is completely spoiled while new viruses are released, ready to attack other bacteria. It should be noticed at this point that among the advantages expected from the therapy is the fact that it focuses on one specific bacteria, while antibiotics also attack autochthonous microbiota. Roughly speaking, it is also believed that viruses are likely to adapt themselves to mutations of their host bacteria.

At a mathematical level, whenever the mobility of the different biological actors is high enough, bacteriophage systems can be modeled by a kind of predator-prey equation. Namely, set  $S_t$  (resp.  $Q_t$ ) for the bacteria (resp. bacteriophages) concentration at time  $t$ . Then a model for the evolution of the couple  $(S, Q)$  is as follows:

$$\begin{cases} dS_t &= [\alpha - k Q_t] S_t dt + \varepsilon S_t dW_t^1 \\ dQ_t &= [d - m Q_t - k Q_t S_t + k b e^{-\mu \zeta} Q_{t-\zeta} S_{t-\zeta}] dt + \varepsilon Q_t dW_t^2, \end{cases} \quad (1)$$

where  $\alpha$  is the reproducing rate of the bacteria and  $k$  is the adsorption rate. In equation (1),  $d$  also stands for the quantity of bacteriophages inoculated per unit of time,  $m$  is their death rate, we denote by  $b$  the number of bacteriophages which is released after replication within the bacteria cell,  $\zeta$  is the delay necessary to the reproduction of bacteriophages (called latency time) and the coefficient  $e^{-\mu\zeta}$  represents an attenuation in the release of bacteriophages (given by the expected number of bacteria cell's deaths during the latency time, where  $\mu$  is the bacteria's death rate). A given initial condition  $(S_0, Q_0)$  is also specified, and the term  $\varepsilon dW_t$  takes into account a small external noise standing for both uncertainties on the measures and the experiment conditions (for similar modeling see e.g. [34]). One should be aware of the fact that the latency time  $\zeta$  (which can be seen as the reproduction time of the bacteriophages within the bacteria) cannot be neglected, and is generally of the same order (about 20mn) as the experiment length (about 60mn).

With this model in hand, our main results in this direction (see [15]) have been the following:

- Quantification of the exponential convergence to a bacteria-free equilibrium of equation (1) when  $d$  is large enough.
- Use of the previous result plus concentration inequalities in order to study the convergence of the noisy system to equilibrium in a reasonable time range.
- Simulation of the stochastic processes at stake in order to observe the convergence to equilibrium.

## 6.5. Convergence of stochastic gene networks

Participants: A. Crudu, A. Debussche, A. Muller, Aurélie, O. Radulescu.

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  $S = 1, \dots, M$ , 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  $X \rightarrow X + \gamma_r$ ,  $\gamma_r \in \mathbb{Z}^M$ .

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  $T_0 = 0$ ,  $T_i = \tau_1 + \dots + \tau_i$ ,  $X$  is constant on  $[T_{i-1}, T_i)$  and has a jump at  $T_i$ . The parameter of  $\tau_i$  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  $X \rightarrow X + \gamma_r$ :  $X(T_i) = X(T_{i-1}) + \gamma_r$ . 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  $\tilde{n}_i = n_i/N$ ,  $\tilde{n}_i$  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 [18] some relevant information on the asymptotic regime  $N \rightarrow \infty$  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.

## 6.6. Inference for Gaussian systems

Participants: F. Baudoin, A. Chronopoulou, S. Cohen, F. Gamboa, Y. Hu, M. Jolis, C. Lacaux, J-M. Loubes, A. Neuenkirch, D. Nualart, C. Ouyang.

(i) *LAN property for fractional Brownian motion.* Local asymptotic normality (LAN) property is a fundamental concept in asymptotic statistics, which gives the asymptotic normality of certain estimators such as the maximum likelihood estimator for instance (see [66] for details on this property). In [11], we focus on the LAN property for the model where we observe a sample of  $n$  observations  $\mathbf{X}_n = (X_1, \dots, X_n)$  of a Gaussian stationary sequence. The sequence  $(X_n)_{n \in \mathbb{N}}$ , whose spectral density  $f_\theta$  is indexed by a parameter  $\theta$ , can admit antipersistence, long memory or short memory and be noninvertible. To be more specific, our main assumption is:

$$f_\theta(x) \sim_{x \rightarrow 0} |x|^{-\alpha(\theta)} L_\theta(x)$$

with  $L_\theta$  a slowly varying function and  $\alpha(\theta) \in (-\infty, 1)$ . We prove the LAN property by studying an asymptotic expansion of the log likelihood and using some results on Toeplitz matrices (see [39], [53]). In particular, our assumptions are fulfilled by fractional Gaussian noises and autoregressive fractionally integrated moving average processes (ARFIMA( $p, d, q$ )). We also obtain the LAN property for fractional Brownian motion.

(ii) *Inference for dynamical systems driven by Gaussian noises.* 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 [64]) is of the form:

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

where  $\theta$  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:

$$X_t^\theta = a + \int_0^t b(\theta; X_u) du + \int_0^t \sigma(\theta; X_u) dB_t, \quad (2)$$

where  $\theta$  enters non linearly in the coefficient, where  $\sigma$  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 2011:

- A better understanding of the underlying rough path structure for equation (2), carried out in [4], [5]. This step allows a proper definition of our equation of interest in a wide range of contexts.
- Gaussian type bounds for equations driven by a fractional Brownian motion, obtained in [9]. This is an important preliminary step for likelihood estimates for stochastic processes.
- Numerical aspects of a maximum likelihood type procedure for an equation of the form (2), expressed in terms of Malliavin calculus tools (see [10]).
- Convergence of a least square type estimator for an equation of the form (2) where the noise enters additively, handled in [14]. This is the first occurrence of a converging estimator for a general coefficient  $b(\theta, \cdot)$ .

## 6.7. Local self-similarity properties and stable or Gaussian random fields

Participants: Hermine Biermé, Jacques Istas, Céline Lacaux, Renaud Marty, Hans-Peter Scheffler.

• Recently, an important class of anisotropic random fields called operator scaling random fields has been studied in [30]. To be more specific, the classical self-similarity property is replaced in [30] by the following operator scaling property:

$$\forall c > 0, (X(c^E x))_{x \in \mathbb{R}^d} \stackrel{(d)}{=} c(X(x))_{x \in \mathbb{R}^d}, \quad \text{where } c^E := \exp(E \ln(c)).$$

The Hölder regularity properties of operator scaling Gaussian or stable harmonizable random fields have been studied in [30] and can be expressed in terms of the matrix  $E$ . In particular, they do not vary along the trajectories, which can be too restrictive for some applications (see our osteoporosis project at the *Application Domains* Section). In order to obtain some anisotropic random fields whose Hölder regularity properties are allowed to vary, we introduce in [1] a local version of the operator scaling property (similar to the local version of the classical self-similarity property defined in [27]). This local property is illustrated in [1], where we also define and study harmonizable multi-operator scaling stable random fields. For such a multi-operator random field, we obtain an accurate upper bound of both the modulus of continuity and global and directional Hölder regularities at any point  $x$ . As expected, the Hölder regularity properties vary along the trajectories.

• In [24], we study the sample paths properties of an anisotropic random field, which is defined as limit of an invariance principle and is of the same type as a multifractional Brownian sheet. Our first aim was to generalize [37], that is to obtain some multifractional random fields indexed by  $\mathbb{R}^d$  with  $d \geq 2$  and to allow Hurst indices to be lower than  $1/2$ . To overcome the problem of the values of the Hurst indices which characterize the limit field, we focus on stationary sequences  $(X_n(H))_{n \in \mathbb{N}}$ , where  $H \in (0, 1)^d$ , defined by an harmonizable representation. Then, our limit field  $S_h$  is defined as the limit of

$$S_h^N = \left\{ \sum_{n_1=1}^{[Nt_1]} \dots \sum_{n_d=1}^{[Nt_d]} \frac{X_n(h_n^N)}{N^{r_n^N}}; t \in [0, +\infty)^d \right\}$$

for some suitable families  $(h_n^N)_{n,N}$  and  $(r_n^N)_{n,N}$ . We then study the sample paths property of this limit field. In particular, we obtain some local self-similarity properties for its increments of order  $k$  and its pointwise global and directional Hölder exponents. We also define (and obtain) some pointwise multi-Hölder exponents which characterize the Hölder property satisfied by the increments of order  $d$  of  $S_h$ .

• We are also interested in self-similar processes indexed by manifolds in [23]. This study is motivated by the fact various spatial data are indexed by a manifold and not by the Euclidean space  $\mathbb{R}^d$  in practical situations such as image analysis.



## 7. Contracts and Grants with Industry

### 7.1. Contracts with Industry

*Start-up project by T. Bastogne:*

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

Status: in incubation.

### 7.2. 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.

*PEPS project (Mathematics-Industry Interactions) led by A. Muller:*

Industrial partner: Sport4Spirit (start-up).

Title: Computation of profit probabilities in sports gambling.

Two Internships involved.

## 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 (Nancy).

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

PhD thesis of M. Lalloué.

*Regional project led 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, led by F. Richard (University of Provence).
- S. Tindel is co-leader the ECRU (Exploration des Chemins Rugueux) ANR project, jointly with M. Gubinelli (University of Paris Dauphine).
- P. Vallois is member of the MASTERIE (Malliavin Stein Random Irregular Equation) ANR project, led by F. Russo (ENSTA, Paris).
- T. Bastogne is leader of the MOCOBIO (MOdeling and COntrol of heterogeneous systems in systems BIOlogy) CNRS-PEPS project.
- T. Bastogne is member of the PDTX (Active Nanoplatforms for Photodynamic Therapy) ANR project, led 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, leaded by M. Barberi-Heyob (Centre de Recherche en Automatique de Nancy, Centre Alexis Vautrin).

## 8.3. European Initiatives

### 8.3.1. Collaborations in European Programs, except FP7

Program: UGR (Université de la Grande Région)

Project acronym: I-DERBI

Project title: I-DERBI

Duration: January 2010 - April 2012

Coordinator: C. Carlberg (Luxembourg)

Other partners: Université du Luxembourg, Université de Liège (Belgium) , Saarland University (Germany)

Abstract: We stand at the brink of a fundamental change in how medicine will be practiced in the next 5-20 years. This change will require the unprecedented integration of biology, medicine, technology and computation as well as societal issues of major importance: ethical, regulatory, public policy, economic, and others. These needs have encouraged the emergence of a biology-based inter-disciplinary study field, systems biology, which focuses on the modeling of complex biological systems. Systems biology covers a large spectrum of applications: biomedicine, bioprocesses engineering, environmental science and pharmaceutical discovery. The ambition of the I-DERBI pilot project is to initiate and develop synergy of education and research in Systems Biology within the Grande Région.

### 8.3.2. Major European Organizations with which Bigs has followed Collaborations

Partner: Universitat Autònoma de Barcelona, Departament de Matemàtiques (Spain).

Subject: Stochastic model for bacteriophage systems.

Partner: TU Kaiserslautern, Department of Mathematics.

Subject: Parameter estimation for differential systems driven by Gaussian processes.

## 8.4. International Initiatives

### 8.4.1. Internships

Yosra Chemli

Subject: Statistical Emulation of High Dimensional Biological Dynamic Models

Institution: Ecole Polytechnique de Tunisie (Tunisia)

Raouf Souabni

Subject: Simulation of the light propagation in biological tissues. Application to interstitial photodynamic therapy.

Institution: Université de Tunis El Manar - Faculté des Sciences (FST) (Tunisia)

## 8.5. Teaching

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

**PhD & HdR:**

PhD : A. Deya, *Etude de systèmes différentiels fractionnaires*, Université de Nancy 1, 18/10/2010, Advisor: S. Tindel.

PhD : R. Keijn, *Modélisation de la croissance d'une tumeur après traitement par radiothérapie*, Université de Nancy 1, 2/12/2011, Advisors: T. Bastogne, P. Vallois.

PhD in progress: R. Bar, *Analyse de données 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. Bibliography

### Publications of the year

#### Articles in International Peer-Reviewed Journal

- [1] H. BIERMÉ, C. LACAUX, H.-P. SCHEFFLER. *Multi-operator Scaling Random Fields*, in "Stochastic Processes and their Applications", 2011, vol. 121, p. 2642-2677, <http://hal.inria.fr/hal-00551707/en>.
- [2] O. COLLIGNON, J.-M. MONNEZ, P. VALLOIS, F. CODREANU, J.-M. RENAUDIN, G. KANNY, M. BRULLIARD, B. BIHAIN, S. JACQUENET, D.-A. MONERET-VAUTRIN. *Discriminant analyses of peanut allergy severity scores*, in "Journal of Applied Statistics", 2011, vol. 38, n<sup>o</sup> 9, p. 1783-1799 [DOI : 10.1080/02664763.2010.529878], <http://hal.archives-ouvertes.fr/hal-00643787/en/>.
- [3] R. KEIJN, T. BASTOGNE, P. VALLOIS. *Multinomial model-based formulations of TCP and NTCP for radiotherapy treatment planning*, in "Journal of Theoretical Biology", June 2011, vol. 279, n<sup>o</sup> 1, p. 55-62 [DOI : 10.1016/J.JTBI.2011.03.025], <http://hal.inria.fr/hal-00588935/en>.
- [4] D. NUALART, S. TINDEL. *A construction of the rough path above fractional Brownian motion using Volterra's representation*, in "Annals of Probability", 2011, p. Vol. 39, no. 3, 1061-1096, 25 pages, <http://hal.inria.fr/hal-00414075/en>.
- [5] S. TINDEL, J. UNTERBERGER. *The rough path associated to the multidimensional analytic fBM with any Hurst parameter*, in "Collectanea Mathematica", 2011, p. Vol 62, no. 2, 197-223, 32 pages., <http://hal.inria.fr/hal-00327355/en>.
- [6] P. VALLOIS, J.-M. MONNEZ. *A novel immunoassay using recombinant allergens simplifies peanut allergy diagnosis*, in "International Archives of Allergy and Immunology", 2011, vol. 154, p. 216-226, <http://hal.inria.fr/hal-00644649/en>.
- [7] P. VALLOIS. *A characterization of Kummer, gamma and beta distributions via Matsumoto-Yor property*, in "Electronic Journal of Probability", 2011, to appear, <http://hal.inria.fr/hal-00644653/en>.

## International Conferences with Proceedings

- [8] T. BASTOGNE, L. TIRAND, J. GRAVIER, D. BECHET, V. MOROSINI, M. PERNOT, C. FROCHOT, A. RICHARD, F. GUILLEMIN, M. BARBERI-HEYOB. *Contributions of experiment designs in photodynamic therapy: photosensitizer design, treatment analysis and optimization.*, in "13th World Congress of the International Photodynamic Association, IPA 2011", Innsbruck, Austria, May 2011, CDROM, Abstract published in *Photodiagnosis and Photodynamic Therapy*, 8(2):137, 2011, <http://hal.inria.fr/hal-00597950/en>.

## Research Reports

- [9] F. BAUDOIN, C. OUYANG, S. TINDEL. *Upper bounds for the density of solutions of stochastic differential equations driven by fractional Brownian motions*, 2011, 27 pages, <http://hal.archives-ouvertes.fr/hal-00639187/en/>.
- [10] A. CHRONOPOULOU, S. TINDEL. *On inference for fractional differential equations*, 2011, 33 pages, 2 figures., <http://hal.archives-ouvertes.fr/hal-00587087/en/>.
- [11] S. COHEN, F. GAMBOA, C. LACAUX, J.-M. LOUBES. *LAN property for some fractional type Brownian motion*, 2011, <http://hal.archives-ouvertes.fr/hal-00638121/en/>.
- [12] Y. HU, S. TINDEL. *Smooth density for some nilpotent rough differential equations*, 2011, Contains a Norris type lemma in the rough paths context. A similar result is to be posted soon by M. Hairer and N. Pillai, <http://hal.archives-ouvertes.fr/hal-00584938/en/>.
- [13] R. KEINJ, T. BASTOGNE, P. VALLOIS. *Tumor growth modeling based on cell lifespans*, 2011.
- [14] A. NEUENKIRCH, S. TINDEL. *A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise*, 2011, 15 pages, <http://hal.archives-ouvertes.fr/hal-00639030/en/>.

## Other Publications

- [15] X. BARDINA, D. BASCOMPTE, C. ROVIRA, S. TINDEL. *A stochastic model for bacteriophage therapies*, 2011, 17 pages, 2 figures, <http://hal.archives-ouvertes.fr/hal-00619307/en/>.
- [16] H. CARDOT, P. CÉNAC, J.-M. MONNEZ. *A fast and recursive algorithm for clustering large datasets with  $k$ -medians*, January 2011, Prépublication IECN 2011/5, <http://hal.archives-ouvertes.fr/hal-00558145/en/>.
- [17] H. CARDOT, P. CÉNAC, J.-M. MONNEZ. *A fast and recursive algorithm for clustering large datasets with  $k$ -medians*, 2011, <http://hal.archives-ouvertes.fr/hal-00644683/en/>.
- [18] A. CRUDU, A. DEBUSSCHE, A. MULLER, O. RADULESCU. *Convergence of stochastic gene networks to hybrid piecewise deterministic processes*, 2011, Prépublication IECN 2011/3; IRMAR 2011/04, <http://hal.archives-ouvertes.fr/hal-00553482/en/>.
- [19] A. DEYA, S. TINDEL. *Malliavin calculus for fractional heat equation*, 2011, Dedicated to David Nualart on occasion of his 60th birthday, <http://hal.archives-ouvertes.fr/hal-00618597/en/>.
- [20] G. DUCHARME, S. FERRIGNO. *An Omnibus Test of Goodness-of-fit for Conditional Distribution Functions with Applications to Regression Models*, February 2011, <http://hal.archives-ouvertes.fr/hal-00641034/en/>.

- [21] S. FERRIGNO, M. MAUMY-BERTRAND, A. MULLER. *Uniform law of the logarithm for the conditional distribution function and application to certainty bands*, July 2011, <http://hal.archives-ouvertes.fr/hal-00641027/en/>.
- [22] Y. HU, M. JOLIS, S. TINDEL. *On Stratonovich and Skorohod stochastic calculus for Gaussian processes*, 2011, 43 pages, <http://hal.archives-ouvertes.fr/hal-00556955/en/>.
- [23] J. ISTAS, C. LACAUX. *On locally self-similar fractional random fields indexed by a manifold*, 2011, <http://hal.archives-ouvertes.fr/hal-00193203/en/>.
- [24] 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/>.

## References in notes

- [25] 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.
- [26] T. BASTOGNE, S. MÉZIÈRES-WANTZ, N. RAMDANI, P. VALLOIS, M. BARBERI-HEYOB. *Identification of pharmacokinetics models in the presence of timing noise*, in "Eur. J. Control", 2008, vol. 14, n<sup>o</sup> 2, p. 149–157, <http://dx.doi.org/10.3166/ejc.14.149-157>.
- [27] A. BENASSI, S. JAFFARD, D. ROUX. *Elliptic Gaussian random processes*, in "Rev. Mat. Iberoamericana", 1997, vol. 13, n<sup>o</sup> 1, p. 19–90.
- [28] J. O. BERGER. *Statistical decision theory and Bayesian analysis*, Springer Series in Statistics, Springer-Verlag, New York, 1993, Corrected reprint of the second (1985) edition.
- [29] H. BIERMÉ, CLAUDE-LAURENT. BENHAMOU, F. RICHARD. *Parametric estimation for Gaussian operator scaling random fields and anisotropy analysis of bone radiographs texture*, in "Proc. MICCAI 12th Int. Conference, London", 2009.
- [30] H. BIERMÉ, C. LACAUX. *Hölder regularity for operator scaling stable random fields*, in "Stochastic Process. Appl.", 2009, vol. 119, n<sup>o</sup> 7, p. 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<sup>o</sup> 11, p. 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<sup>o</sup> 2, p. 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<sup>o</sup> 1, 4.

- [34] M. CARLETTI. *Mean-square stability of a stochastic model for bacteriophage infection with time delays*, in "Math. Biosci.", 2007, vol. 210, n<sup>o</sup> 2, p. 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, p. 1–53.
- [36] S. COHEN, C. LACAU, M. LEDOUX. *A general framework for simulation of fractional fields*, in "Stochastic Process. Appl.", 2008, vol. 118, n<sup>o</sup> 9, p. 1489–1517, <http://dx.doi.org/10.1016/j.spa.2007.09.008>.
- [37] S. COHEN, R. MARTY. *Invariance principle, multifractional Gaussian processes and long-range dependence*, in "Ann. Inst. Henri Poincaré Probab. Stat.", 2008, vol. 44, n<sup>o</sup> 3, p. 475–489.
- [38] 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/>.
- [39] R. DAHLHAUS. *Efficient parameter estimation for self-similar processes*, in "Ann. Statist.", 1989, vol. 17, n<sup>o</sup> 4, p. 1749–1766.
- [40] A. DEYA, S. TINDEL. *Rough Volterra equations. I. The algebraic integration setting*, in "Stoch. Dyn.", 2009, vol. 9, n<sup>o</sup> 3, p. 437–477, <http://dx.doi.org/10.1142/S0219493709002737>.
- [41] 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/>.
- [42] 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<sup>o</sup> 3, p. 297–338, With comments and a reply by the authors, <http://dx.doi.org/10.1198/073500102288618397>.
- [43] 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<sup>o</sup> 5, p. 313–316, <http://dx.doi.org/10.1016/j.crma.2005.07.003>.
- [44] 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<sup>o</sup> 1-2, p. 83–86, <http://dx.doi.org/10.1016/j.crma.2007.11.026>.
- [45] M. GUBINELLI, S. TINDEL. *Rough evolution equations*, in "Ann. Probab.", 2010, vol. 38, n<sup>o</sup> 1, p. 1–75, <http://dx.doi.org/10.1214/08-AOP437>.
- [46] S. C. KOU. *Stochastic Networks in Nanoscale Biophysics*, in "Journal of the American Statistical Association", 2008, vol. 103, n<sup>o</sup> 483, p. 961-975, <http://pubs.amstat.org/doi/abs/10.1198/016214507000001021>.
- [47] Y. A. KUTOYANTS. *Statistical inference for ergodic diffusion processes*, Springer Series in Statistics, Springer-Verlag London Ltd., London, 2004.

- [48] C. LACAUX. *Real Harmonizable Multifractional Lévy Motions*, in "Ann. Inst. Poincaré.", 2004, vol. 40, n<sup>o</sup> 3, p. 259–277.
- [49] C. LACAUX. *Series representation and simulation of multifractional Lévy motions*, in "Adv. in Appl. Probab.", 2004, vol. 36, n<sup>o</sup> 1, p. 171–197, <http://dx.doi.org/10.1239/aap/1077134469>.
- [50] 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.
- [51] 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/>.
- [52] 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.
- [53] O. LIEBERMAN, R. ROSEMARIN, J. ROUSSEAU. *Asymptotic Theory for Maximum Likelihood Estimation of the Memory Parameter in Stationary Gaussian Processes*, in "Econometric Theory", 2011, p. 1–14, to appear [DOI : 10.1017/S0266466611000399], <http://www.esaim-cocv.org/action/displayAbstract?fromPage=online&aid=8376039&fulltextType=RA&fileId=S0266466611000399>.
- [54] H. LIERO. *Testing homoscedasticity in nonparametric regression*, in "Journal of Nonparametric Statistics", 2003, vol. 15, n<sup>o</sup> 1, p. 31-51, <http://www.tandfonline.com/doi/abs/10.1080/10485250306038>.
- [55] T. LYONS, Z. QIAN. *System control and rough paths*, Oxford mathematical monographs, Clarendon Press, 2002, <http://books.google.com/books?id=H9fRQNIngZYC>.
- [56] 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)..
- [57] J.-M. MONNEZ. *Approximation stochastique en analyse factorielle multiple*, in "Ann. I.S.U.P.", 2006, vol. 50, n<sup>o</sup> 3, p. 27–45.
- [58] J.-M. MONNEZ. *Stochastic approximation of the factors of a generalized canonical correlation analysis*, in "Statist. Probab. Lett.", 2008, vol. 78, n<sup>o</sup> 14, p. 2210–2216, <http://dx.doi.org/10.1016/j.spl.2008.01.088>.
- [59] J.-M. MONNEZ. *Convergence d'un processus d'approximation stochastique en analyse factorielle*, in "Publ. Inst. Statist. Univ. Paris", 1994, vol. 38, n<sup>o</sup> 1, p. 37–55.
- [60] 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<sup>o</sup> 3, p. 327–342, <http://dx.doi.org/10.1016/j.anihpb.2005.04.005>.
- [61] A. NEUENKIRCH, I. NOURDIN, A. RÖSSLER, S. TINDEL. *Trees and asymptotic expansions for fractional stochastic differential equations*, in "Ann. Inst. Henri Poincaré Probab. Stat.", 2009, vol. 45, n<sup>o</sup> 1, p. 157–174, <http://dx.doi.org/10.1214/07-AIHP159>.

- [62] 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](http://dx.doi.org/10.1007/978-3-540-71189-6_7).
- [63] S. TINDEL. *Quenched large deviation principle for the overlap of a p-spins system*, in "J. Statist. Phys.", 2003, vol. 110, n<sup>o</sup> 1-2, p. 51–72, <http://dx.doi.org/10.1023/A:1021062510565>.
- [64] C. A. TUDOR, F. G. VIENS. *Statistical aspects of the fractional stochastic calculus*, in "Ann. Statist.", 2007, vol. 35, n<sup>o</sup> 3, p. 1183–1212, <http://dx.doi.org/10.1214/009053606000001541>.
- [65] É. WALTER, L. PRONZATO. *Identification of parametric models*, Communications and Control Engineering Series, Springer-Verlag, Berlin, 1997, From experimental data, Translated from the 1994 French original and revised by the authors, with the help of John Norton.
- [66] A. W. VAN DER VAART. *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 1998, vol. 3.