RESEARCH CENTRE

Inria Centre at Université de Lorraine

IN PARTNERSHIP WITH:

Université de Lorraine, CNRS

2024

ACTIVITY REPORT

Project-Team

SIMBA

Statistical Inference and Modeling for Biological Applications

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

DOMAIN

Digital Health, Biology and Earth

THEME

Modeling and Control for Life Sciences

Inria

Contents

Project-Team SIMBA

Creation of the Project-Team: 2024 January 01

Keywords

Computer sciences and digital sciences

A3.1. – Data

A3.2. – Knowledge

A3.2.3. – Inference

A3.3. – Data and knowledge analysis

A3.3.1. – On-line analytical processing

A3.3.2. – Data mining

A3.3.3. – Big data analysis

A3.4.1. – Supervised learning

A3.4.2. – Unsupervised learning

A3.4.5. – Bayesian methods

A3.4.7. – Kernel methods

A6.1. – Methods in mathematical modeling

A6.1.1. – Continuous Modeling (PDE, ODE)

A6.1.2. – Stochastic Modeling

A6.1.3. – Discrete Modeling (multi-agent, people centered)

A6.1.4. – Multiscale modeling

A6.2. – Scientific computing, Numerical Analysis & Optimization

A6.2.2. – Numerical probability

A6.2.3. – Probabilistic methods

A6.2.4. – Statistical methods

Other research topics and application domains

B1. – Life sciences

B1.1. – Biology

B1.1.2. – Molecular and cellular biology

B1.1.4. – Genetics and genomics

B1.1.6. – Evolutionnary biology

B1.1.8. – Mathematical biology

B1.1.10. – Systems and synthetic biology

B1.1.11. – Plant Biology

B2.2. – Physiology and diseases

B2.2.1. – Cardiovascular and respiratory diseases

B2.2.3. – Cancer

B2.3. – Epidemiology

B2.4. – Therapies

B2.4.2. – Drug resistance

1 Team members, visitors, external collaborators

Research Scientists

- Nicolas Champagnat [Team leader, INRIA, Senior Researcher]
- Coralie Fritsch [INRIA, Researcher]
- Ulysse Herbach [INRIA, Researcher]
- Bruno Scherrer [INRIA, Researcher, until Jun 2024]
- Edouard Strickler [CNRS, Researcher]

Faculty Members

- Sandie Ferrigno [UL, Associate Professor]
- Anne Gegout Petit [UL, Professor]
- Jean-Marie Monnez [UL, Emeritus]
- Aurélie Muller [UL, Associate Professor]
- Pierre Vallois [UL, Emeritus]
- Denis Villemonais [UL, Associate Professor]
- Sophie Wantz-Mézières [UL, Associate Professor]

PhD Students

- Sophie Baland [UL]
- Virgile Brodu [UL]
- Mathilde Gaillard [INRIA]
- Vincent Kagan [UL]
- Anouk Rago [UL, ATER, from Oct 2024]
- Vidhi Vidhi [INRIA, from Oct 2024]

Technical Staff

- Nathaniel Seyler [INRIA, Engineer, until Sep 2024]

Interns and Apprentices

- Pascaline Kouda [UL, Intern, from Apr 2024 until Sep 2024]
- Vidhi Vidhi [INRIA, Intern, from Apr 2024 until Sep 2024]

Administrative Assistant

- Emmanuelle Deschamps [INRIA]

2 Overall objectives

SIMBA is a joint team of Inria, CNRS and University of Lorraine, within the Institut Élie Cartan de Lorraine (IECL), UMR 7502 CNRS-UL, laboratory in mathematics, of which Inria is a strong partner. The team is composed of applied mathematicians whose research interests mainly concern probability and statistics with applications in biology and medicine.

Many fundamental questions and applications in medicine and biology critically rely on our capacity to construct, estimate, analyse and simulate complex mathematical models. They can aim at building predictions and decision making processes upon heterogeneous, noisy, incomplete or inconsistent data, at improving the understanding of complex phenomena involving several interacting subsystems that can only be calibrated separately, or at supplying a priori information on phenomena that cannot be reproduced in laboratory experiments or for which data are too costly to collect. In the past years, these models have been gradually refined, taking into account more interactions or dependencies between different components, more heterogeneity between subsystems and a wider range of time or space scales. Along with this gradual complexification, the importance of stochasticity has been recognized as fundamental in many biological or medical studies, either to take into account intrinsic randomness in biological processes, to evaluate confidence in a model's parameters or predictions, or to take into account randomness or uncertainties in environmental conditions. In parallel, the specificities of biomedical data, which are typically high dimensional, heterogeneous, correlated and with few observations, and their gradual increase pose new statistical challenges in terms of classification, prediction, variable selection or streaming data analysis.

Our expertise gathers a large spectrum of mathematical domains, ranging from statistics to stochastic modeling and analysis. We share a common experience and dedication to interactions with other sciences (biology, medicine) and interested parties (physicians, clinical researchers in the medical domain, start-ups). Our application domains are medicine, epidemiology, systems biology, ecology and evolution. The specificity of our group is our capacity to use a broad range of tools to answer practical statistical, modeling and analytical questions posed by collaborators from biology or medicine.

We believe that an interdisciplinary approach is crucial to answer questions posed by biologists or physicians, but can also bring original mathematical questions requiring the development of new theoretical tools which may apply to a broader range of application domains. Our ambition is therefore both to follow a bottom-up research program, where we tackle practical modeling or statistical questions posed by practitioners, and a top-down approach where we develop new mathematical tools to study general models and questions from biology and medicine. Part of our work is purely mathematical, but always motivated by biological applications.

Statistical and stochastic modeling are central to our project. The range of mathematical models we study is large, but they all share common features: we are mainly interested in dynamical models of biological populations with interactions (in a general sense). In our models, populations can be composed of individuals (in ecology, evolution or epidemiology), cells (e.g. bacteria in ecology or tumor cells in medicine), genes or proteins (in systems biology), or populations in the statistical sense (e.g. patients in epidemiology). Interactions or coupling can occur within a population or between species (e.g. in evolution), between cells (e.g. in oncology), between genes or proteins (intracellular networks), with an environment (e.g. in tumor growth, where environmental conditions are linked to medical treatment, or in ecology) or between high-dimensional statistical variables (e.g. clustering and variable selection in epidemiology). In biology and medicine, models were primarily developed to understand phenomena at fixed time and space scales. Today, more and more models aim at studying phenomena at large scales resulting from delicate coupling and interactions between scales. Descriptions at small scales are typically high dimensional and often involve stochastic features, or a combination of stochastic and deterministic features (hybrid models). Our research project aims at studying such complex systems using analytical tools to construct: 1. coarse-grained representations of small scale features (through averaging, ergodic limits, homogenization, mean field models...); 2. specific numerical methods, possibly based on coarse-grained representations, to efficiently bridge the gap between different time and space scales; 3. appropriate statistical inference or learning tools often based on limited or partial observation, in order to make predictions.

3 Research program

The research challenges we present in this section are mainly theoretical or methodological. All of them are motivated by biological or medical applications and provide a wide methodological toolbox that can be combined to answer biological or medical questions.

3.1 Stochastic modeling for health

The modeling issues we address in medicine aim at understanding fundamental mechanisms of cancer development, understanding how cells make decisions through gene expression and bringing new insights on the evolution of telomere length distribution with age and across a population. Telomeres are non-coding regions of repetitive nucleotide sequences located at each end of chromosomes. In human, They shorten at each cell division and it is known that short telomere lengths are statistically linked to age related diseases. These transversal applications are described in Section 4.

Population dynamics of tumor cells have been modeled in numerous works, either by deterministic models (*ordinary differential equations* or ODEs and *partial differential equations* or PDEs), or stochastic ones, either discrete (*birth and death* or *branching processes, individual-based models* in infinite dimension) or continuous (*stochastic differential equations* or SDEs). Branching processes or individual-based models are also fundamental tools to study telomere length dynamics. Concerning gene networks, in addition to classical models (e.g. Gaussian graphical models or deterministic systems with external noise), we develop new models (see Section 3.2.3) based on PDMPs (*piecewise-deterministic Markov processes*). Our team gathers experts of all these classes of models, both from the analytical, statistical and numerical simulation points of view. In most applications we have in mind, we need to combine within-cell dynamics (such as telomere shortening or gene expression) with cell population dynamics, leading to multiscale and/or multicomponent hybrid models. Multicomponent models are also ubiquitous in medicine when one takes into account latent variables such as the genealogical tree of mutations within a tumor when only observations of clonal population sizes, that is sizes of populations of cells sharing the same genetic material, are available.

3.1.1 Links between macroscopic models and stochastic microscopic processes

For multiscale models, it is often relevant to distinguish what could be called the *microscopic level* (i.e. the level of single individuals), the *macroscopic level* (i.e. the level of large population densities) or the intermediate *mesoscopic level* (i.e. a level of population densities, but where demographic stochasticity cannot be neglected). Microscopic models can be for example stochastic, individual-based models, mesoscopic models can be SDEs and macroscopic models are usually models of population densities, such as ODEs, PDEs or PDMPs. Many biological questions are stated at the macroscopic level, and the main modeling issue lies in the appropriate way to incorporate meso- or microscopic features in the description of the macroscopic scale.

Typically, individual-based models are models where the population state at time t is described as a counting measure $\nu_t = \sum_{i=1}^{N_t} \delta_{x_i(t)}$ where N_t is the number of individuals alive at time t, and $x_i(t)$ is the characteristic of the i-th individual (e.g. phenotype, mass, size, length of telomeres, age...), belonging to some set $\mathscr{X}$. The dynamics is strongly dependent on the precise phenomenon to be modeled, but in the simplest cases, $(\nu_t)_{t\geq 0}$ is a Markov jump process on point measures on $\mathscr{X}$ whose infinitesimal generator has the following form:

$$L\phi(\nu) = \iint_{\mathscr{X}^2} (\phi(\nu+\delta_y)-\phi(\nu))\, b(x,\nu;dy)\nu(dx) + \int_{\mathscr{X}} (\phi(\nu-\delta_x)-\phi(\nu))\, d(x,\nu)\nu(dx), \tag{1}$$

where $b(x,\nu;dy)$ is the infinitesimal birth rate of an individual y from an individual x in the population ν and $d(x,\nu)$ is the death rate of an individual x in the population ν.

We have developed a strong expertise in the derivation of simplified macroscopic models from complex microscopic effects when there is a strong separation between time and/or space scales. Mathematically, this requires us to encode the scale separation through scaling parameters and to solve an asymptotic analysis problem, which can be averaging (slow-fast dynamics, singular perturbation) [47, 2], homogeneization [39, 54], concentration limits [42, 45, 46] or more generally parameter scaling problems

where the scaling parameter has a biological meaning [44, 40]. For individual-based models, the simplest parameter scaling that can be considered is a large population asymptotic encoded with a parameter K, where the state of the population is modified as $\nu_t^K = \frac{1}{K}\sum_{i=1}^{N_t}\delta_{x_i(t)}$ and the generator (1) as

$$L^K\phi(\nu^K) = K\iint_{\mathscr{X}^2}\left[\phi\left(\nu+\frac{1}{K}\delta_y\right)-\phi(\nu)\right]b(x,\nu^K;dy)\nu^K(dx) + K\int_{\mathscr{X}}\left[\phi\left(\nu-\frac{1}{K}\delta_x\right)-\phi(\nu)\right]d(x,\nu^K)\nu^K(dx), \tag{2}$$

leading to mean-field macroscopic models when $K\to\infty$ [44]. More complex scalings can involve rare or small mutations combined with long-time scalings, complex local interactions as in [39] or multiscale phenomena (e.g. within cell dynamics combined with the dynamics of populations of cells, like for telomeres). Each problem usually requires new methodological development.

3.1.2 Numerical analysis

Sometimes, it is not possible to construct a simplified macroscopic description of complex, multiscale biological phenomena. In such cases, we need to rely on numerical simulations. More generally, numerical simulations are very helpful to supply *a priori* information on phenomena that are difficult to reproduce in laboratory experiments or on large-scale models that involve several interacting elements.

Despite increasing computing facilities, existing numerical schemes are rarely adapted to individual-based models. Better numerical approaches based on a deeper understanding of their multiscale features would significantly reduce computational costs and yield reliable error estimates. This is one of our motivations in studying model reduction through asymptotic analysis as described above. We propose to design numerical schemes taking into account local extinction or local deterministic approximation, or develop hybrid methods relying on duality methods based e.g. on the Poisson representation of birth and death processes, which allows to switch between microscopic and mesoscopic models when the population size crosses a threshold.

3.2 Analysis of biological and medical data

Modern medicine is turning to highly personalized approaches, and a major challenge is to design and develop a new generation of techniques to assist prevention, diagnosis, prognosis and therapy. A major difficulty is the integration and exploitation of data which are often high-dimensional, heterogeneous, incomplete or inconsistent, to build predictions and decision making processes. The main part of our research in statistical learning aims to develop operational tools for the analysis of data from our collaborations with biologists and physicians. Another part of our project builds upon the sophisticated models described in Section 3.1, for which specific inference tools need to be designed.

3.2.1 Statistical learning, regression

We want to develop methodologies that take into account the specificities of biological data: they are often high dimensional and correlated (for instance multiomics data, such as genome, proteome or transcriptome) but with few observations (patients) and with missing data. Variable selection is of particular interest to study the link between an outcome (the occurence of an illness for instance) with covariates or to infer partial correlations between several variables for instance to study quantitative microbiome data. For instance, in variable selection, we propose to study the theoretical guarantees of our methods [32, 59] and to extend them to other models of dependencies such as mixtures of quantitative or qualitative covariates with dependencies between the covariates. In regression, we wish to tackle the challenge not only to select the covariates related to an event (illness, death...), but also to understand which configuration of these covariates triggers its occurrence.

We also develop goodness-of-fit tests to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are “directional” in that they detect departures from a given assumption of the model. Other tests are “omnibus” in that they assess whether a model fits a dataset on all its assumptions. We focus on the task of choosing the structural part of the regression function because it contains easily interpretable information about the studied relationship. Among the large number of existing tests, we consider nonparametric tests which are all based on generalizations of the Cramér-von Mises statistic. To perform these goodness-of-fit tests, we develop the R package `cvmgof` [31,

1], an easy-to-use tool which allows practitioners to compare the implemented tests. In the future, we plan to enrich the `cvmgof` package with tests concerning the other assumptions of a regression model such as the functional form of the variance or the additivity of the random error term, using directional tests such as [56]. Another perspective is to develop a similar tool for other statistical models widely used in biostatistics such as generalized linear models.

3.2.2 Signal or online data analysis

We develop tools for the analysis of online data, which are now very frequent in the health domain (recording of cardiac signals, physiological measurements via connected objects...). The purpose is either the update of estimation parameters for models with sequential arrival of new data, or the online detection of change-points in temporal signals. For the first purpose, stochastic algorithms are essential tools [41], which allow one to approximate eigenvectors in a stepwise manner [71, 70, 72]. We focus on several incremental procedures for regression and data analysis like PCA (Principal Component Analysis) [73] and linear and logistic regressions on standardised data in order to avoid numerical explosion [64]. Our aim is to apply these results to other methods related to PCA, such as multiple factorial analyses, partial PCA, and to learning (classification, regression, event scores).

Change-point detection is of particular interest in e-health, for example for detecting changes in the health status of the elderly or people at risk of a disease. We plan to develop tools for online change-point detection using score-based CUSUM statistics. Our aim is to use the beginning of the signal to build simulation-based thresholds that have to be crossed for the detection of the change-point. We also want to propose new stopping rules adapted to the characteristics of the signal. These questions may be relevant for a large part of the applications we shall develop with physicians.

3.2.3 Network inference

Inferring networks from data is often a crucial step for understanding biological interactions, such as regulatory links between genes within individual cells [60, 38] or communication relationships between bacteria within a population.

Concerning gene regulatory networks, we are interested in so-called single-cell data, where mRNA levels are measured individually in many cells rather than being population-averaged, revealing the intrinsically stochastic "transcriptional bursting" phenomenon. In previous work [60], we described a promising strategy in which the network inference problem is seen as a calibration procedure for a PDMP model that is able to fit real single-cell data [10]. In the simplest version of this model, the state of each gene $i \in \{1,\dots,n\}$ at time t is described by its promoter state $E_i(t) \in \{0,1\}$, the quantity of transcribed RNA $Y_i(t) \in \mathbb{R}_+$ and of translated proteins $X_i(t) \in \mathbb{R}_+$, where

$$\begin{gathered} E_i(t) \text{ jumps from 0 to 1 at rate } k_{\mathrm{on},i}(X(t)) \text{ and from 1 to 0 at rate } k_{\mathrm{off},i}, \\ \dot{Y}_i(t) = s_{0,i}E_i(t) - d_{0,i}Y_i(t), \qquad \dot{X}_i(t) = s_{1,i}Y_i(t) - d_{1,i}X_i(t). \end{gathered} \tag{3}$$

Interactions between genes in the network are encoded in functions $k_{\mathrm{on},i}(x) \ll k_{\mathrm{off},i}$ which correspond to burst frequencies. We plan to develop dynamical models that fully exploit the particular time structure of the data: as cells have to be killed for measurements, such data do not consist of cell trajectories but rather independent samples of time-varying multivariate distributions. As the number of genes can be large ($\approx 2 \times 10^5$ in humans), we will also investigate variational methods, which generalize the usual expectation-maximization (EM) algorithm, as a relevant way to make the inference procedure both robust and scalable.

We also develop methods to infer gene networks from dynamic gene expression data within a collaboration with CHRU Strasbourg. The goal here is to design new models and inference methods adapted to the prediction of the outcome of biological intervention experiments such as "gene knock-down", in order to identify therapeutic targets that could be experimentally studied.

3.2.4 Inference for stochastic processes

In our collaborations with practitioners, we are led to infer specific quantities of interest for stochastic dynamical models of various classes, such as the speed of telomere shortening using multitype branching

processes; the growth rate of clonal populations and their phylogenetic tree using heterogeneous population growth models in cancer; links between gene expressions using PDMPs or the speed of propagation of fungi using stochastic models with latent variables given by the solution of a partial differential equation (see our main applications in Section 4). For all these sophisticated stochastic processes, statistical inference raises many open, difficult questions. In particular, we plan to develop inference tools for general classes of models such as PDMP, bifurcating Markov processes or branching processes. More generally, the inference of dynamical models relies strongly on their long-time behavior, for which the team has a strong expertise (see Section 3.3). Most often in biology, models rely on latent variables, that may follow complex dynamics as in the examples above. The inference of this kind of models requires us to develop efficient Bayesian algorithms, EM like algorithms and variational methods.

3.3 Stochastic modeling for ecology and evolution

In ecology, we are specifically interested in theoretical challenges in conservation biology (the branch of biology dealing with extinction and survival of species) and in the response of ecosystems to environmental perturbations such as climate change, anthropization or niche construction. In evolution, we have a strong expertise in the study of the long term evolution of biological populations using approximate models based on various biological assumptions. Although this application domain is different from the one of Section 3.1, the questions we address are close from the mathematical point of view.

3.3.1 Links between macroscopic and microscopic eco-evolutionary models

We develop here similar ideas as in Section 3.1.1, but the mathematical questions are of different nature for models of tumor growth in medicine than for population models in ecology or evolution, because in the first case one wants to capture transitory behavior (e.g., in growing populations for tumor growth), whereas in the second case, one usually models long term evolution assuming that the ecological dynamics is in a stationary (quasi-equilibrium) state and that evolution acts slowly.

In addition to the various classes of models described in Section 3.1.1, which are also relevant for ecology and evolution, some other types of models are well-developed in these domains, such as Dawson–Watanabe processes [51], Fleming–Viot processes [57] or a particular class of PDMPs called *switched dynamical systems* used to model abruptly changing environments. The general goal of designing macroscopic models from complex multiscale models through parameter scalings extends to these new classes of models. For example, Fleming-Viot processes appear as the fast dynamics in the long term evolution of biological populations under assumptions of small mutations and large population [2].

In ecology, our motivations are to highlight the biological assumptions underlying different classes of macroscopic models, or to take into account at the macroscopic scale complex local interactions between individuals. In evolutionary biology, the time scales involved are so long that it is hard to observe experimentally evolutionary phenomena such as diversification. Mathematical analysis of models is therefore of great importance, e.g. to construct approximate models which allow us to the predict long term evolution of biological populations.

3.3.2 Adaptive dynamics: concentration limits

Asymptotic analysis is particularly useful in the branch of evolutionary biology called *adaptive dynamics*. This biological theory which studies the interplay between ecological interactions and long term evolution was developed in the mid 90's [67, 52] and provides theoretical ecologists with useful tools to deduce evolutionary patterns from ecological parameters (directional evolution through *canonical equations* and diversification through *evolutionary branching*).

So far, two mathematical approaches to justify, study and improve these tools were developed: a deterministic one, based on PDE models [53, 74, 66] and a probabilistic one, based on individual-based models [42, 46]. Both approaches are concentration limits, aiming to construct approximate models where population densities are replaced by Dirac masses representing coexisting sub-populations. They both can be seen as particular parameter scalings on individual-based models of the form (2), combining large population scaling with scalings of rare mutations and/or small mutations.

However, the adaptive dynamics toolbox and the existing mathematical approaches are criticized because they are based on unrealistic biological assumptions on the scales involved in the process [79, 75]. Mathematical analysis is needed both to quantify the underlying assumptions on scales and to propose alternative models based on more realistic assumptions. For example, we plan to design PDE models of adaptive dynamics allowing for local extinctions of populations [61, 68]. Similarly, the actual occurrence of evolutionary branching in sexual populations is debated [79] and our goal is to shed light on these questions using asymptotic analysis from individual-based models.

3.3.3 Population dynamics with absorption and quasi-stationary distributions

In conservation biology, it is fundamental to quantify the chances of survival of species in a given environment. In addition, the observed biological populations are intrinsically conditioned to be non-extinct, which introduces an observation bias that is rarely taken into account. For a given model of population dynamics, it is therefore important to develop tools which allow us to study the population size before extinction and to quantify its extinction probability in a given time window. When the population size is stable during long time intervals before extinction, it can be described by a quasi-stationary distribution (QSD), defined as a stationary distribution conditionally on non-extinction. The QSD also allows us to quantify the population extinction rate.

Our research program on this topic builds on recent works of the team [49, 4], where we developed probabilistic criteria for the large time convergence of conditional distributions of stochastic population processes, that proved to apply to a wide range of stochastic processes. These works open many perspectives. We focus here on the methodological ones and will mention some numerical issues in Section 3.3.4 below. A first question is to obtain criteria for the convergence of conditional distributions for weaker distances than in [49, 4]: instead of the total variation distance, we study convergence in Wasserstein distances. This is particularly relevant for PDMPs or for infinite dimensional processes such as individual-based models, where coupling properties are not strong enough to expect convergence in total variation. We also want to study other questions related to QSDs: Can we characterize the speed of convergence of conditional distributions to a QSD? Is it possible to study the path to extinction of a stochastic population process, in order to characterize the parameters improving their survival (e.g. for protected species) or their extinction (e.g. for pests in agronomy)? What can be said about the genealogy of populations before extinction?

3.3.4 Numerical analysis

The challenges detailed here are related to those described in Section 3.1.2. Ecological or evolutionary models pose specific numerical problems that we want to tackle. For example, in the numerical simulation of individual-based models like (1), when the population size is small, randomness cannot be neglected and exact algorithms have to be used; when the population size is large, such algorithms become too costly and one would like to take advantage of deterministic approximations like those we developed in [43, 58]. The error analysis of such hybrid numerical schemes is difficult [30, 63] and the case of spatially or trait-structured individual-based models is still largely open.

We are also interested in numerical methods to approximate QSDs, among which the most developed are particle methods [78, 48] and stochastic algorithms such as self-interacting processes [33, 35]. The analysis of these algorithms relies on long-time analysis of particle or nonlinear systems and our research project on QSDs detailed above will be of great help. In particular, Wasserstein distances are known to be well adapted to the study of convergence of particle systems thanks to their tensorization properties. We also plan to develop new numerical approaches based on the stationary distribution of approximating processes, such as those obtained by central limit theorems for stochastic processes [50].

4 Application domains

We have described in the last section our theoretical expertise and several mathematical challenges. However, our main motivation comes from our interactions with biologists, physicians or clinical researchers. Most often, these interactions involve several of the methodological tools developed above, that need to be combined for precise biological goals. Our strategy is to establish collaborations with few groups

of biologists and physicians, that allow us to tackle ambitious, long term projects. In this section, we illustrate the different domains of application mentioned in the previous section by describing several ongoing pluridisciplinary projects that involve large subgroups of the team.

4.1 Tumor growth and heterogeneity

4.1.1 Reconstruction of tumor heterogeneity

Targeted therapies represent a real advance in the treatment of patients with cancer. Most of these therapies are kinase inhibitors and require precise analysis of tumor DNA mutations to ensure the absence of primary resistance. Indeed, tumors are often genetically heterogeneous with the presence of many subclones, but they release "circulating" cell-free DNA (cfDNA) that can be directly extracted from basic blood samples: as measurement sensitivity improves, such *liquid biopsies* increasingly appear as a mirror of tumor heterogeneity. In this context, we are taking a promising statistical approach to analyze longitudinal cfDNA data, with the purpose of gaining a deeper understanding of the mechanism by which resistance develops in individual patients. While addressing the standard problem of reconstructing the associated phylogenetic tree, this approach also describes the production of cfDNA from the temporal dynamics of cells, in order to best exploit the longitudinal structure of the data. This is a project in collaboration with physicians Jean-Louis Merlin (Institut de Cancérologie de Lorraine), Alexandre Harlé (Institut de Cancérologie de Lorraine) and Erwan Pencreac'h (CHRU Strasbourg). We currently also have another ongoing project on a tumor heterogeneity reconstruction for chronic lymphocytic leukemia in collaborationb with Laurent Vallat (CHRU Strasbourg).

4.1.2 Evolution of low-grade gliomas

We have an ongoing collaboration with the Centre de Recherche en Automatique de Nancy CRAN (Jean-Marie Moureaux) and neuro-oncologist and surgeon from CHRU Nancy (Luc Taillandier, Fabien Rech) about diffuse low-grade gliomas (DLGG). These are slow-growing tumors that are often asymptomatic for a long period of time. They progress to a higher grade, resulting in the patient's death. The current treatment strategy aims to surgically reduce tumor volume as soon as possible. As DLGGs infiltrate functional areas, surgery is performed in an awake state, with active patient participation, while electrical brain stimulations are done, to identify functional structures. At CHRU Nancy, surgeries are filmed, but the patient's responses are recorded manually. We aim to develop an automatic tool to detect, analyse and register the patients responses. We use deep learning algorithms for motion and speech detection. In perspective, the fine anomalies that can be identified are likely to be correlated with the patient's short- and long-term cognitive outcome.

4.2 Telomeres

Telomeres are non-coding regions of repetitive nucleotide sequences located at each end of a chromosome. They protect the end of the chromosome from deterioration or from fusion with neighboring chromosomes, ensuring the integrity of genetic material over cell divisions. At each cell division, telomeres loose a short fragment, a phenomenon often called the 'end replication problem'. When its telomeres are too short, the cell stops dividing and enters a senescence phase. In human, it is known that short telomere lengths are statistically linked to age related diseases [37].

In an ongoing collaboration with Athanase Benetos (CHRU of Nancy) and Simon Toupance (CHRU of Nancy), we study the telomere length distribution in a human body, its relation with the patient's phenotype, its evolution with age and across generations. Our contribution to the project is to bring competences ranging from theoretical probabilities to applied statistics, including modelling and numerical simulation of stochastic models and analysis of distributional data of telomeres length. A first goal is to provide physicians with additional medical statistics to better understand the health state of a patient using its telomere length distribution, based on our observation that the shape of the telomere length distribution is stable accross ages of an individual, leading to the concept of *telomere signature* [9]. We plan to study larger cohorts of individuals with medical records and parental relationships to construct an equivalence relationship between shapes of distributions and to develop health scores for patients allosing to assess the risk of particular diseases. A second goal is to bring new insights for the description

and modelling of the evolution of telomere length distribution with age and across a population. For this, we study a stochastic branching model of the telomere length into a given tissue of the form of (1) where $x_i(t)$ is the vector of lengths of telomeres of cell i. We also plan to study models for the evolution of the telomere length accross a population of individuals and on evolutionary time scales, where $x_i(t)$ is the telomere lengths in gametes of individual i. This project requires advanced mathematical tools related to the theory of branching processes and of non-conservative semi-groups. We also plan to tackle parameter estimation questions for these models.

In another ongoing project with Marie-Noëlle Simon (CRCM, Aix-Marseille Université), we study modeling and inference of the different mechanisms of telomere shortening or elongation in survivor cells of yeast *Saccharomyces cerevisiae* for which telomerase is inactivated. Telomerase is an enzyme that is active in normal yeasts and compensates telomere shortening due to the end replication problem. When telomerase in inactivated, most yeasts undergo replicative senescence, except a few ones called survivor cells, which are able to develop alternative telomere elongation mechanisms. The goal of the project is to develop a full model of the evolution of survivor cells, in particular by estimating the rates and sizes of abrupt telomere shortening or lengthening in survivor cells.

4.3 Gene networks and single-cell data

4.3.1 Modeling gene expression at single-cell level

Gene expression in cells has long been only observable through averaged quantities over cell populations. The development of single-cell transcriptomics has enabled gene expression to be measured in individual cells: it turns out that even for an isogenic population located in a homogeneous medium, molecular variability can be large. An average description is therefore not sufficient to account for fundamental phenomena such as cell differentiation. Recently, a view emerged that the dynamics governing the switching of cells from one differentiation state to another could be characterized by a peak in gene expression variability at the point of fate commitment [77]. We are continuing on this path, working on the link between PDMP models and notions of entropy and epigenetic landscape.

4.3.2 Transcriptional bursting in regulatory networks

Working in the active field of single-cell dynamics and gene regulatory networks provides opportunities to interact with biologists such as Olivier Gandrillon from ENS Lyon [77, 12], and potentially also physicians such as Erwan Pencreac'h in CHRU Strasbourg. Related to this last point, we notice that the biological literature increasingly highlights gene regulatory networks as playing an important role (independent of genetic mutations) in the acquisition of resistance to cancer treatments: hence this topic might become soon also relevant to the application area of oncology.

4.3.3 Prediction and identification of therapeutic targets for chronic lymphocytic leukemia

In an ongoing collaboration with Laurent Vallat (CHRU Strasbourg), we develop new models and inference methods for gene regulation networks allowing to make prediction of biological intervention experiments (such as gene knock-down). Inference is performed on gene expression data from cells of patients suffering from different forms of chronic lymphocytic leukemia. The goal is to use prediction to identify therapeutic targets which could be knocked-down to reduce the cells' proliferation. Biological experiments will then be performed by Laurent Vallat and his group to assess the therapeutic potential of the new targets.

4.4 Chalara

The Chalara project [24] is a team project with Benoît Marçais (INRAE Champenoux) and Marie Grosdidier (INRAE Avignon). Chalara is an ash disease that arrived in France 12 years ago through the Grand Est and has been spreading throughout France ever since. The disease spreads by means of fungus spores that are deposited on the leaves of trees during summer, fall at the foot of the trees during fall and give rise to fungi that release new spores that spread the following summer. Affected trees show signs of decline (defoliation, canker...) that can lead to their death.

The objective is to model the spread of chalara and to study and quantify the potential underlying environmental effects, such as humidity or high temperatures. We use a hybrid model where spores spread is based on reaction-diffusion PDEs and other steps of the disease cycle are stochastic. The project requires skills in modeling, statistical methodology and simulations. Future work will consist in choosing between several diffusion or dispersion models, for example using Bayesian model averaging, which gives weights to different models to assess which one is better suited for prediction.

5 Highlights of the year

Note : Readers are advised that the Institute does not endorse the text in the "Highlights of the year" section, which is the sole responsibility of the team leader.

From the scientific point of view, the team had the following milestones in 2024:

- The team is involved is the target project DyLT of PEPR Maths VivES on the dynamics of telomere length distribution. This project is coordinated by Nicolas Champagnat and Athanase Benetos (CHRU Nancy).
- Anne Gégout-Petit was chair of the program committee of ENBIS 2024 (Leuven, 15-19 september)
- Denis Villemonais was hired Professor at the Université de Strasbourg in September.

From the institutional point of view, Inria management has imposed on the institute a new Contract of Objectives, Means and Performance (COMP) with the French government, for the period 2024–2028. It presages major changes for Inria, regarding both its missions and the way it operates. These changes, whose precise nature and impact on the staff are unclear, should become effective as soon as 2025 but have not been the subject of any consultation, and inasmuch as the collaboration of Inria's staff is necessary to turn this disruption into a successful change, we are concerned that the top management has remained deaf to several votes and petitions opposing these policies.

The multiplication of new missions and priorities, particularly those related to the "program agency" or oriented towards defence applications, pushes the research carried out at Inria in the background. The constraints induced by this COMP will restrain the independence of scientists and teams, as well as their freedom to select research topics and collaborators.

Our main concern with the COMP lies with the following items:

- Placement of Inria in a "zone à régime restrictif" (ZRR).
- Restriction of international and industrial collaborations to partners chosen by the institute's management, with no clear indication of the rules.
- Individual financial incentives for researchers involved in strategic partnerships, whose topics are steered by the program agency.
- Priority given to "dual" research with both military and civilian applications, materialised by tighter links with the Ministry of Defence.

6 New software, platforms, open data

6.1 New software

6.1.1 cvmgof

Keywords: Regression, Test, Estimators

Scientific Description: Many goodness-of-fit tests have been developed to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are "directional" in that they detect departures from a given assumption of the model. Other tests are "global" (or "omnibus") in that they assess whether a model fits a dataset on all its assumptions. cvmgof focuses on the task of choosing the structural part of the regression function because it contains easily interpretable

information about the studied relationship. It implements 2 nonparametric "directional" tests and one nonparametric "global" test, all based on generalizations of the Cramer-von Mises statistic.

Functional Description: cvmgof is an R library devoted to Cramer-von Mises goodness-of-fit tests. It implements three nonparametric statistical methods based on Cramer-von Mises statistics to estimate and test a regression model.

URL: https://cran.r-project.org/web/packages/cvmgof/index.html

Publication: hal-03101612

Contact: Romain Azais

Participants: Sandie Ferrigno, Marie-Jose Martinez, Romain Azais

6.1.2 Harissa

Name: Hartree approximation for inference along with a stochastic simulation algorithm

Keywords: Gene regulatory networks, Reverse engineering, Molecular simulation

Functional Description: Harissa is a Python package for both inference and simulation of gene regulatory networks, based on stochastic gene expression with transcriptional bursting. It was implemented in the context of a mechanistic approach to gene regulatory network inference from single-cell data.

URL: https://github.com/ulysseherbach/harissa

Publications: hal-04208601, hal-04071033, hal-01646910

Contact: Ulysse Herbach

6.1.3 MultiRNAflow

Name: An R package for the analysis of RNAseq raw counts with multiple biological conditions and time points

Keywords: RNA-seq, Gene regulatory networks, Integrated data analysis, Complex experimental design, Multiple temporal and biological conditions, Differential expression

Functional Description: The R package MultiRNAflow provides an easy to use unified framework allowing to make both unsupervised and supervised analysis (differential expression analysis) for RNAseq datasets with an arbitrary number of biological conditions and time points. In particular, this package makes a deep downstream analysis of differential expression information, e.g. identifying temporal patterns across biological conditions and differentially expresses genes which are specific to a biological condition for each time.

Release Contributions: First version

URL: https://bioconductor.org/packages/release/bioc/html/MultiRNAflow.html

Contact: Nicolas Champagnat

Participants: Rodolphe Loubaton, Nicolas Champagnat, Pierre Vallois, Laurent Vallat

Partner: CHRU de Strasbourg

6.1.4 quantCurves

Keyword: Statistical modeling

Functional Description: Non-parametric methods as local normal regression, polynomial local regression and penalized cubic B-splines regression are used to estimate quantiles curves.

URL: https://cran.r-project.org/web/packages/quantCurves/quantCurves.pdf

Contact: Sandie Ferrigno

7 New results

7.1 Stochastic modeling for health, ecology and evolution

7.1.1 Quasi-stationary distributions

Participants: Nicolas Champagnat, Denis Villemonais.

External collaborators: SIMSMART project-team, MATHERIALS project-team, CERMICS

Together with Tony Lelièvre, Mouad Ramil and Julien Reygner, we studied in [22] kinetic SDEs with low regularity coefficients in the setting recently introduced in [76]. For the solutions to such equations, we first prove a Harnack inequality. Using the abstract approach of [4], this inequality then allows us to prove, under a Lyapunov condition, the existence and uniqueness (in a suitable class of measures) of a quasi-stationary distribution in cylindrical domains of the phase space. We finally exhibit two settings in which the Lyapunov condition holds true: general kinetic SDEs in domains which are bounded in position, and, following [65], Langevin processes with a non-conservative force and a suitable growth condition on the force.

7.1.2 Multi-type bisexual branching process

Participants: Coralie Fritsch, Denis Villemonais.

The asexual multi-type Galton-Watson branching processes as well as the single-type bisexual processes have been studied in the literature. In particular, survival condition of the processes are well known in both cases. However, until now, the multi-type bisexual branching processes have only been studied in very specific situations and no general mathematical description has been established yet. In [11], we studied general multi-type bisexual branching processes with superadditive mating function. We exhibited a necessary and sufficient condition for almost sure extinction, we proved a law of large numbers for our model and we studied the long-time convergence of the rescaled process. In [25] we study the quasi-stationary behavior of this model in the subcritical case.

7.1.3 Dispersal Induced Growth

Participants: Edouard Strickler.

External collaborators: Michel Benaïm (Université de Neuchâtel, Suisse), Claude Lobry (Université de Nice) and Tewfik Sari (Inrae Montpellier)

In this collaboration, we exhaustively studied the phenomenon of *Dispersal Induced Growth* (DIG), a term coined by Katriel in [62]. In a population spreading across a finite number of patches, individuals in

patch i undergoes a time-varying growth rate r_i which is such that, in the absence of migration between the patches, the population will eventually become extinct in each patch. There is Dispersal Induced Growth when adding migration between this patches lead the whole population to grow and to survive. In the papers [36, 34] we developed several tools in order to understand when the DIG phenomenon can happen. In case where the graph of migration is connected at each time, then DIG happens if and only if $\chi > 0$ where χ is the mean growth rate of an idealized habitat which would have the maximal growth rate at each time. In [17, 18], we consider the case where the migration is time dependent, the graph of migration is globally connected but not necessarily connected at each time. In this situation, we exhibit example where $\chi > 0$ but there is no DIG. We also give example of Dispersal Induced Decay, meaning that, in absence of migration, the population will survive in each patch, while adding migration may lead the whole population to extinction. In [36, 34, 17], the study is performed via the sign of a Lyapunov exponent, leading to necessary and sufficient condition for DIG. In [18], we developed a new approach using elementary minorization tools. This give us sufficient conditions for DIG that apply in a much larger class of models.

7.1.4 Asymptotic expansion of Lyapunov exponent

Participants: Edouard Strickler.

External collaborators: Pierre Monmarché (Paris Sorbonne Université) and Sebastian Schreiber (DEEP Lab, UC Davis, USA)

The long term behaviour of a large class of population models can be deduced from the sign of a Lyapunov exponent, linked to the growth rate of the population when rare. However, these exponents are in general not computable, and knowing the sign of a Lyapunov exponent can be non tractable. This is why, in this collaboration, we give some asymptotic expansion of Lyapunov exponents for linear cooperative systems. In details, we studied population model given by randomly or periodically switched linear ODEs, associated to a finite set of matrices $\{A_1, \dots, A_n\}$. It was already known that when switching quickly between these matrices, the Lypaunov exponent is close to the principal eigenvalue of the matrix $\bar{A}$, which is a mean of the matrices A_i, while when switching slowly, the Lyapunov exponent is close to a mean of the principal eigenvalues of the matrices A_i. In [69] and [26], we precise these asymptotic behaviours, by specifying the first order term in the Taylor expansion of the Lyapunov exponent seen as a function of the switching rate, when this rate goes to 0 or infinity. We apply in particular our results to the DIG phenomenon to prove that, contrary to the symmetric case (studied in [62]), when the migration is not symmetric between the patches, the Lyapunov exponent is not necessarily a monotone function of the switching rate.

7.1.5 Large population asymptotics for an epidemic model in switching environment

Participants: Edouard Strickler.

External collaborators: Adrien Prodhomme (Université de Tours)

In [16], we look at a process modelling the evolution of a disease in a population divided into d distinct groups. Infections between individuals and recovery times are random variables with an exponential distribution, whose parameters can change randomly as a function of changes in the underlying environment. The population is finite, of size N. In this case, the disease will almost surely disappear from the population. However, when N tends to infinity, the typical extinction time T_N of the disease will also tend to infinity. Our aim is to understand the asymptotic behaviour of T_N. We show that this behaviour is strongly dependent on the behaviour of the limiting process, which is a system of ordinary differential equations in a random environment (PDMP). More precisely, this limit process has a Lyapunov exponent

Λ such that if $\Lambda \leq 0$, the disease will almost surely disappear from the population, whereas if $\Lambda > 0$, the law of the process will converge towards a unique stationary distribution μ^* describing the lasting establishment of the disease in the population. For the extinction time of the disease when the population is of size N, we show that

1. If $\Lambda < 0$, then T_N is of order $\ln(N)$;
2. Si $\Lambda > 0$, then T_N is of order N^p for some $p > 0$ that we can characterize.

In addition, the quasi-stationary distribution of the birth-death process in a random environment converges as N tends to infinity towards the Dirac mass in 0 in the case where $\Lambda < 0$, and towards μ^* in the case where $\Lambda > 0$.

7.1.6 Towards a stochastic theory of niche construction

Participants: Nicolas Champagnat, Coralie Fritsch, Edouard Strickler.

External collaborators: Universidad de Valparaiso, Pontificia Universidad Catolica de Chile, Santa Fe Institute, Universidad de Santiago de Chile, INRAE Montpellier

In collaboration with Rolando Rebolledo, Pablo Marquet, Leonardo Videla, Cristobal Quininao and Nicolas Zalduendo-Vidal, we are working on the modeling of the eco-evolutionary process of niche construction, by which a species or an ecological community is able to modify its environment in such a way that the induced adaptation enhances survival of the species or the community.

We are currently working on two articles. The first one deals with a model of sublinear growth of populations. This research started during the visit to Chile of Nicolas Champagnat in March. We study the properties of extinction and survival of population models with sublinear growth rate parameterized by an exponent θ. We point out that this family of models, recently proposed to fit various population growths, may have important flaws depending on the value of the parameter, which make them unrealistic for ecological modeling. We both study birth-death processes and diffusion processes.

We also currently work on a toy-model of niche construction based on birth-death processes of d interacting (sub)species immersed in an environment which is influenced by the population state (the so-called niche construction) and which evolves on a slower time-scale. Under the above hypotheses, extinction and/or re-emergence of negligible species on long time scales can be observed. We prove that the joint dynamics of the logarithm of the species sizes and the environment undergo a piecewise deterministic Markov process, which can be approximated by an explicit dynamical system in the limit of large populations.

7.1.7 Telomere

Participants: Sophie Baland, Nicolas Champagnat, Coralie Fritsch, Denis Villemonais.

External collaborators: CHRU Nancy, CRCM Aix-Marseille Université

In most cells, with each cell division, telomeres shorten due to the so-called end replication problem, which can lead to replicative senescence and a variety of age-related diseases. On the other hand, in certain cells, the presence of the enzyme telomerase can lead to the lengthening of telomeres, which may delay or prevent the onset of such diseases but can also increase the risk of cancer. We developed, in collaboration with reserachers of CHRU Nancy, a stochastic representation of this biological model, which takes into account multiple chromosomes per cell, the effect of telomerase, different cell types and the dependence of the distribution of telomere length on the dynamics of the process. We study theoretical properties of this model, including its long-term behaviour. In addition, we investigate numerically the

impact of the model parameters on biologically relevant quantities, such as the Hayflick limit and the Malthusian parameter of the population of cells [19].

In another ongoing project with Marie-Noëlle Simon (CRCM, Aix-Marseille Université), we start to study modeling and inference of the different mechanisms of telomere shortening or elongation in survivor cells of yeast *Saccharomyces cerevisiae* for which telomerase is inactivated. The telomerase enzyme is active in normal yeasts and compensates telomere shortening due to the end replication problem. When telomerase is inactivated, most yeasts undergo replicative senescence, except a few ones called survivor cells, which are able to develop alternative telomere elongation mechanisms. The goal of the project is to develop a full model of the evolution of survivor cells, in particular by estimating the rates and sizes of abrupt telomere shortening or lengthening in survivor cells.

7.1.8 Individual-based model of allometric relationships in Ecology

Participants: Virgile Brodu, Nicolas Champagnat, Coralie Fritsch.

External collaborator: Sylvain Billiard (Université de Lille)

In [20], we design a stochastic individual-based model structured in energy, for single species consuming an external resource, where populations are characterized by a fixed positive energy at birth. The resource is maintained at a fixed amount, so we benefit from a branching property at the population level. We focus on individual trajectories, constructed as PDMP with random jumps modelling births and deaths in the population and a continuous and deterministic evolution of energy between jumps. We are interested in the case where metabolic (i.e. energy loss for maintenance), growth, birth and death rates depend on the individual energy over time, and follow allometric scalings (i.e. power laws). Our goal is to determine in a bottom-up approach what are the possible allometric coefficients (i.e. exponents of these power laws) under elementary and ecologically relevant constraints, for our model to be valid for the whole spectrum of possible body sizes. We show in particular that assuming an allometric coefficient for metabolism strongly constrains the range of possible values for the allometric coefficients for birth, death and growth rates.

7.2 Analysis of biological and medical data

7.2.1 Prediction of silencing experiments on gene networks for chronic lymphocytic leukemia

Participants: Nicolas Champagnat, Anne Gégout-Petit, Anouk Rago, Pierre Vallois.

External collaborators: CHRU Strasbourg

In this collaboration with the group of Laurent Vallat in CHRU Strasbourg, we work on the inference of dynamical gene networks from RNAseq data of chronic lymphocytic leukemia. The goal is to infer a model of gene expression allowing to predict gene expression in cells where the expression of specific genes is knocked-down (e.g. using siRNA), in order to select the knock-down experiments which are more likely to reduce cell proliferation. We expect the selected genes to provide new therapeutic targets for the treatment of chronic lymphocytic leukemia. This year, we have developed a new method of prediction of the effect of gene knock-down, based on penalized regression models and on a priori biological information about gene regulators from the ENCODE database. To this aim, we developed an original two-step optimization procedure to solve a specific form of weighted LASSO that we call semi-LASSO [21]. We also developed the package MultiRNAflow (see Section 6.1.3) for the exploratory analysis and the detection of differentially expressed genes from temporal gene expression datasets with several biological conditions [13].

7.2.2 Quantifying and predicting the evolution of clonal heterogeneity in chronic lymphocytic leukemia

Participants: Nicolas Champagnat, Coralie Fritsch, Ulysse Herbach, Pierre Vallois, Vidhi Vidhi.

External collaborators: CHRU Strasbourg

The development of targeted therapies has allowed considerable progress in the treatment of many cancers, but their efficacy is dependent on intra-tumor heterogeneity. In lymphomas and leukemias, the identification of gene alterations by high-throughput sequencing allows the characterization of this heterogeneity. In these hemopathies, the initial leukemic clone has a unique immune repertoire corresponding to a specific VDJ gene sequence encoding the antigen receptor. The occurrence of additional mutations in VDJ genes may be responsible for the emergence of subclones with increased antigen receptor reactivity further complicating the clonal heterogeneity of these hemopathies. However, this second level of clonal heterogeneity and its evolution remain poorly characterized and is not considered in the management of these cancers.

In collaboration with the group of Laurent Vallat in CHRU Strasbourg, we aim to develop a mathematical model for the evolution of the two levels of clonal heterogeneity in leukemia, allowing to characterize their evolution from longitudinal bulk sequencing data of VDJ and cancer genes mutations using a Bayesian approach. This year, we developed a model of the phylogenetic tree of squences of VDJ genes based on the weighted uniform distribution over trees, with weights related to the Hamming distance which counts the number of mutations between sequences. This model allows us to represent with a graph the mots probable phylogenetic trees and the uncertainties of certain parts of the tree. We are currently working on the development of variational methods of inference of the tree in the context of unobserved VDJ sequences.

7.2.3 Stochastic modeling of a gene regulatory network driving B cell development in germinal centers

Participants: Ulysse Herbach.

External collaborators: ENS de Lyon, Université Paris Cité

Germinal centers (GCs) are the key histological structures of the adaptive immune system, responsible for the development and selection of B cells producing high-affinity antibodies against antigens. Due to their level of complexity, unexpected malfunctioning may lead to a range of pathologies, including various malignant formations. One promising way to improve the understanding of malignant transformation is to study the underlying gene regulatory networks (GRNs) associated with cell development and differentiation. Evaluation and inference of the GRN structure from gene expression data is a challenging task in systems biology. Recent achievements in single-cell (SC) transcriptomics allow the generation of SC gene expression data, which can be used to sharpen the knowledge on GRN structure.

In order to understand whether a particular network of three key gene regulators (BCL6, IRF4, BLIMP1), influenced by two external stimuli signals (surface receptors BCR and CD40), is able to describe GC B cell differentiation, we used a stochastic model to fit SC transcriptomic data from a human lymphoid organ dataset [12]. The model is defined mathematically as a piecewise-deterministic Markov process [60]. We showed that after parameter tuning, the model qualitatively recapitulates mRNA distributions corresponding to GC and plasmablast stages of B cell differentiation. Thus, the model can assist in validating the GRN structure and, in the future, could lead to better understanding of the different types of dysfunction of the regulatory mechanisms.

7.2.4 Deciphering rind color heterogeneity of smear-ripened Munster cheese and its Association with microbiota

Participants: Sandie Ferrigno.

External collaborators: LIBio - Laboratoire d'Ingénierie des Biomolécules (F. Borges, A. Petit)

In this collaboration with the LIBio - Laboratoire d'Ingénierie des Biomolécules of Lorraine University, we work on the quality of cheese. Color is one of the first criteria to assess the quality of cheese. However, very limited data are available on the color heterogeneity of the rind and its relationship with microbial community structure. In this study, the color of a wide range of smear-ripened Munster cheeses from various origins was monitored during storage by photographic imaging and data analysis in the CIELAB color space using luminance, chroma, and hue angle as descriptors. Different levels of inter- and intra-cheese heterogeneity were observed. The most heterogeneous Munster cheeses were the darkest with orange-red colors. The most homogeneous were the brightest with yellow-orange colors. K-means clustering revealed three clusters distinguished by their color heterogeneity. Color analysis coupled with metabarcoding showed that rinds with heterogeneous color exhibited higher microbial diversity associated with important changes in their microbial community structure during storage. In addition, intra-cheese community fluctuations were associated with heterogeneity in rind color. The species *Glutamicibacter arilaitensis* and *Psychrobacter nivimaris/piscatorii* were found to be positively associated with the presence of undesirable brown patches. This study highlights the close relationship between the heterogeneity of the cheese rind and its microbiota. An article [14] has been published on this work.

7.2.5 Harnessing ecological niche modeling of Listeria monocytogenes for biopreservation system engineering

Participants: Sandie Ferrigno.

External collaborators: LIBio - Laboratoire d'Ingénierie des Biomolécules (F.Borges)

In this collaboration with the LIBio - Laboratoire d'Ingénierie des Biomolécules of Lorraine University, we work on the presence of pathogens in food. To reduce this presence, hurdle technology, which is based on the use of a combination of several preservative methods, is used by food business operators. Among the multiple available hurdles, biopreservation consists of using microorganisms as protective cultures and/or their metabolites to improve the microbial quality of food. This study explores the potential of ecological niche modeling to guide the selection of biopreservation candidates. A luminescent strain of Listeria monocytogenes was utilized in a multivariate high-throughput competition assay. The resulting data were analyzed using two parallel methods: k-means clustering and Response Surface Modeling. An article is nearly finalized.

7.2.6 Nonparametric estimation

Participants: Sandie Ferrigno.

External collaborators: R. Azais (MOSAIC INRIA Team, ENS Lyon), M.-J. Martinez (LJK-Grenoble University) and INSERM (EDEN Cohort)

Many goodness-of-fit tests have been developed to assess the different assumptions of a (possibly heteroscedastic) regression model. Most of them are 'directional' in that they detect departures from a given assumption of the model. Other tests are 'global' (or 'omnibus') in that they assess whether

a model fits a dataset on all its assumptions. We focus on the task of choosing the structural part of the regression and the variance functions because they contain easily interpretable informations about the studied relationship. We consider two nonparametric 'directional' tests and one nonparametric 'global' test, all based on generalizations of the Cramér-vonMises statistic. To perform these goodness-of-fit tests, we have developed the R package cvmgof, an easy-to-use tool for practitioners, available from the Comprehensive R Archive Network (CRAN). The package was updated in 2022 (this is its third version)(6.1.1). This latest version currently allows testing the regression function in the model. In 2024, we worked to enrich the package by allowing the user to test the homoskedasticity/heteroskedasticity of the model. This new version will be submitted to CRAN in 2025 and an associated article is currently being written.

In epidemiology, reference or standard curves are required to study fetal and child development. Values that lie outside the limits of these reference curves may indicate the presence of a disorder. Some classical empirical, parametric and semi-parametric methods, such as polynomial regression and LMS methods, are usually used to construct these curves. However, these classical methods are built upon restrictive assumptions on estimated curves. We used alternative nonparametric methods such as Nadaraya-Watson kernel estimation, local polynomial estimation, B-splines or cubic splines to construct these curves. Different methods to choose smoothing parameters or choice of knots for the different types of nonparametric estimation are also proposed. To fit these curves, the R package quantCurves, an easy-to-use tool for practitioners, has been developped in 2022 (see Section 6.1.4) and also a graphical interface to enable intuitive visualization of the results of the package. We presented different applications of this package to the EDEN cohort (INSERM) during the CMStatistics congress in London in December 2024 [27].

7.2.7 Online Big Data Analysis and Online Learning

Participants: Jean-Marie Monnez.

A tool for analyzing streaming data is stochastic approximation introduced by Robbins and Monro in 1951, that can be used for example to estimate online parameters of a regression function [55] or centers of clusters in unsupervised classification [41]. Another type of stochastic approximation processes was introduced by Benzécri in 1969 for estimating eigenvectors and eigenvalues of the unknown Q-symmetric expectation of a random matrix A using independent observations of A. In all these processes, it is assumed that independent observations of the random matrix are oberved and that one or a mini-batch of observations per step are taken into account. We are interested in the study of cases where we can't have independent observations and we define processes where at each step all the observations up to this step are taken into account without storing them. Experiments we have conducted show that this second type of process generally converges faster than the first type.

In the article [8], we establish an almost sure convergence theorem of an extension of the stochastic approximation process of Oja for estimating eigenvectors of the unknown Q-symmetric expectation B of a random matrix, under a correlation model between the incoming random matrices. This theorem generalizes previous theorems and extends them to the case where the metric Q is unknown and estimated online in parallel. We suggest constructing processes using past and current observations at each step without storing them. We prove the almost sure convergence of specific processes to corresponding eigenvalues. We apply these results to streaming PCA of a random vector Z, when a mini-batch of observations of Z is used at each step or all the observations up to the current step. We deal with the case of streaming generalized canonical correlation analysis, interpreted as a PCA with a metric estimated estimated online in parallel.

Given a non-decreasing sequence of closed convex subsets $(K_n)_n$ in a separable real Hilbert space H, we study the convergence of a stochastic approximation process in H with correlated observations, projected at step n on K_n, extending processes of the Robbins-Monro type. We established theorems of almost sure convergence and in quadratic mean of this process and apply them to streaming multidimensional linear regression, using at each step a mini-batch of data or all the data up to this step, and to dynamic generalized linear models when the parameter varies with time. This work will be submitted

soon.

7.2.8 Chalara

Participants: Anne Gégout-Petit, Coralie Fritsch.

External collaborators: INRAE Champenoux, INRAE Avignon.

In [24], we have developed a mechanistic statistical model that describes the spread of the chalara disease of ash trees caused by fungi *H. fraxineus*. It takes into account climate (summer temperature and spring rainfall), pathogen population dynamics (foliar infection, Allee effect induced by limited sexual partner encounters) and host density. We fitted this model using available disease reports. We estimated the parameters of our model, first identifying the appropriate ranges for the parameters, which led to a model reduction, and then using an adaptive multiple importance sampling algorithm for fitting. The model reproduces well the propagation observed in France over the last 20 years. In particular, it predicts the absence of disease impact in the south-east of the country and its weak development in the Garonne valley in south-western France. Summer temperature is the factor with the highest overall effect on disease spread, and explains the limited impact in southern France. Among the different temperature indices tested, the number of summer days with temperatures above 28°C gave the best qualitative behavior and the best fit. In contrast, the Allee effect and the heterogeneity of spring precipitation did not strongly affect the overall expansion of *H. fraxineus* in France and could be neglected in the modeling process. The model can be used to infer the average annual dispersal of *H. fraxineus* in France.

7.2.9 Diffuse Low-grade Gliomas

Participants: Sophie Wantz-Mézières.

External collaborators: CRAN, CHRU Nancy.

The therapeutic management of patients with DLGG is based on monitoring progress through regular MRIs, usually through the reconstructed volume of the tumor (after semiautomatic delineation). But this seems not sufficient. Up to now, the diffuse nature of this kind of tumors has been observed but is not well measured. We designed a new MRI-based variable, the ESVR (ExtraSphere Volume Ratio) to quantify the DLGG brain infiltration and discriminate patterns of patients. A machine learning approach allows us to detect that patients'age and ESVR at diagnosis seem to play an important role, as well as the well-known anatomopathology results. This result is about to be submitted, and has been partly presented in the conference ENBIS-2024.

8 Partnerships and cooperations

Participants: Sophie Baland, Virgil Brodu, Nicolas Champagnat, Coralie Fritsch, Mathilde Gaillard, Anne Gegout Petit, Ulysse Herbach, Jean-Marie Monnez, Anouk Rago, Edouard Strickler, Pierre Vallois, Vidhi Vidhi, Denis Villemonais, Sophie Wantz-Mézières.

8.1 International initiatives

8.1.1 Inria associate team not involved in an IIL or an international program

MAGO

Title Modelling and analysis for growth-fragmentation processes

Duration 2022-2024

Coordinator Denis Villemonais

Partners
- Inria team SIMBA (C. Fritsch, D. Villemonais)
- University College London (F.X. Briol, O. Key, A. Watson)

Objective Growth-fragmentation (GF) refers to a collection of mathematical models in which objects – classically, biological cells – slowly gather mass over time, and fragment suddenly into multiple, smaller offspring. These models may be used to represent a range of biological processes, in which an individual reproduces by fission into two or more new individuals, such as the evolution of plasmids in bacteria populations and protein polymerisation. It is crucial to understand the long-term behaviour of GF processes so that they can be used to build algorithms to simulate real-world processes and estimate quantities such as the growth rate of the system, the steady state behaviour, and the fragmentation rate and kernel, allowing scientists to gain a better understanding of the behaviour of these complex systems. In this project, we aim to combine probabilistic and statistical tools to study these processes. In particular, we will employ methods from branching processes, quasi-stationary distributions and interacting particle systems to study their long-term behaviour and develop numerical simulations. Further, we will develop likelihood-free methods to estimate the model parameters, followed by goodness-of-fit tests to analyse the strength of these methods when working with real data.

AStoNiche

Title Towards a stochastic theory of niche construction

Duration 2024-2026

Coordinators Nicolas Champagnat and Rolando Rebolledo

Partners
- Inria team SIMBA (N. Champagnat, C. Fritsch, E. Strickler, N. Zalduendo-Vidal)
- Universidad de Valparaiso (R. Rebolledo, N. Rivera)
- Pontificia Universidad Catolica de Chile (P. Marquet, C. Quininao)
- Universidad de Santiago de Chile (L. Videla)

Objectives We aim to provide a general stochastic formulation of the niche construction process. In particular, we want to take into account the feedbacks of species on their environment, and the evolutionary aspects that follow. This requires to deal with different time-scales (ecological, niche construction, evolutionary...) and to keep track of non-extinct traits that may be positively selected after niche construction. We plan to use mean-field stochastic individual-based model and branching processes and consider appropriate parameter scalings.

8.2 International research visitors

8.2.1 Visits of international scientists

Cristobal Quininao

Status Researcher

Institution of origin: Pontificia Universidad Catolica de Chile

Country: Chile

Dates: October 28 - October 31

Context of the visit: Associate team AStoNiche

Mobility program/type of mobility: Research stay

Leonardo Videla

Status researcher

Institution of origin: Universidad de Santiago de Chile

Country: Chile

Dates: October 28 - October 31

Context of the visit: Associate team AStoNiche

Mobility program/type of mobility: Research stay

Tobias Hurth

Status researcher

Institution of origin: Université de Neuchâtel

Country: Switzerland

Dates: November 25 – November 28

Context of the visit: Collaboration on conditioned Brownian motion

Mobility program/type of mobility: Research stay

8.2.2 Visits to international teams

Denis Villemonais

Visited institution: Université de Neuchatel

Country: Switzerland

Dates: February 26 - March 1

Context of the visit: Collaboration with Julien Reygnier, Lucas Journel, Tony Lelièvre (CERMICS and team Inria MATHERIAL) and with Michel Benaïm (Univ. Neuchatel) on quasi-stationary distributions and interacting particles systems.

Mobility program/type of mobility: Research stay

Nicolas Champagnat

Visited institution: Pontificia Universidad Católica de Chile and Universidad de Valparaiso

Country: Chile

Dates: March 16 - March 27

Context of the visit: Collaboration with Pablo Marquet and Rolando Rebolledo on niche construction within the associate team AStoNiche

Mobility program/type of mobility: Research stay

Edouard Strickler

Visited institution: University of Amsterdam

Country: Netherland

Dates: August 26 - August 29

Context of the visit: Collaboration with Maximilian Engel and Tobias Hurth on conditioned Brownian motion

Mobility program/type of mobility: Research stay

Coralie Fritsch & Denis Villemonais

Visited institution: University College London

Country: UK

Dates: October 28 - October 31

Context of the visit: Collaboration with Alex Watson and Emma Horton on the growth-coagulation-fragmentation processes in the framework of the associate team MAGO.

Mobility program/type of mobility: Research stay

8.3 European initiatives

8.3.1 H2020 projects

N. Champagnat is scientific collaborator of the ERC SINGER (AdG 101054787) on Stochastic dynamics of sINgle cells, coordinated by S. Méléard (Ecole Polytechnique). He is involved in the research axes "From stochastic processes to singular Hamilton-Jacobi equations" and "Lineages and time reversed trajectories" of this project.

8.4 National initiatives

Projects coordinated by the team:

INSERM funding Project Predi-CLL, ITMO Physics, Mathematics applied to Cancer (from October 2023): "Quantifying and predicting the evolution of clonal heterogeneity in chronic lymphocytic leukemia". Funding organisms: ITMO Cancer, ITMO Technologies pour la santé de l'alliance nationale pour les sciences de la vie et de la santé (AVIESAN), INCa. Partners: Inria and IECL (Institut Élie Cartan de Lorraine) and CHRU Strasbourg. Leader: N. Champagnat. Participants: C. Fritsch, U. Herbach, P. Vallois, V. Vidhi, D. Villemonais.

France 2030 funding PEPR Exploratoire Maths-VivES, (starting in spring 2024), target project DyLT (Dynamics of Telomere Length) on "Influence of telomere length dynamics and environmental conditions on biological and clinical aspects of aging"'. Funding organisms: ANR. Partners: Inria Nancy and Saclay, Institut Élie Cartan de Lorraine (Nancy), CHRU Nancy, Centre de Recherche en Cancérologie de Marseille and Institut de recherche sur le cancer et le vieillissement (Nice). Coordinators: N. Champagnat and A. Benetos (CHRU Nancy). Participants: C. Fritsch, A. Gégout-Petit, D. Villemonais, S. Baland.

Projects in which the team participates:

France 2030 funding PEPR Santé Numérique (started in July 2023), project AI4scMed (Multiscale AI for single-cell-based precision medicine), team involved in WP3: "Regulatory network inference: from dynamical models to logical models". Funding organisms: ANR. Partners: Inria, Inserm, CNRS. Coordinator: F. Picard (CNRS, ENS Lyon). Participants: M. Gaillard, U. Herbach.

VEOLIA funding Chair "Modélisation Mathématique et Biodiversité" between VEOLIA, Ecole Polytechnique, Museum National d'Histoire Naturelle and Fondation X. Coordinator: S. Méléard. Participants: V. Brodu, N. Champagnat, C. Fritsch, D. Villemonais.

ANR funding ANR JCJC project CRESCENDO (inCRease physical Exercise and Sport to Combat ENDOmetriosis, AAPG 2022). Coordinator: G. Escriva-Boulley (LISEC, Université de Haute-Alsace). Participant: U. Herbach.

CNRS funding GDR 720 IASIS. Leader: C. Richard. Participant: S. Wantz-Mézières.

CNRS funding GDR Réseau Thématique MathSAV. Leader: F. Crauste. Participants: N. Champagnat, C. Fritsch, U. Herbach, A. Rago.

Univ. Lorraine funding FHU CARTAGE (Fédération Hospitalo Universitaire Cardial and ARTerial AGEing). Leader: Pr. A. Benetos. Participants: J.-M. Monnez, A. Gégout-Petit.

8.5 Regional initiatives

A. Gégout-Petit is one the two PIs of the interdisciplinary program "Life Travel" of the I-Site "Lorraine Université d'Excellence" on life trajectories and longevity (in maturation).

9 Dissemination

Participants: Sophie Baland, Virgil Brodu, Nicolas Champagnat, Sandie Ferrigno, Coralie Fritsch, Mathilde Gaillard, Anne Gegout Petit, Ulysse Herbach, Vincent Kagan, Jean-Marie Monnez, Anouk Rago, Nathaniel Seyler, Edouard Strickler, Pierre Vallois, Vidhi Vidhi, Denis Villemonais, Sophie Wantz-Mézières.

9.1 Promoting scientific activities

9.1.1 Scientific events: organisation committee, scientific committee

- N. Champagnat is member of the scientific committee of the "12ème Biennale Française des Mathématiques Appliquées et Industrielles" (SMAI 2025) congress.
- N. Champagnat organizes an invited session at the 44th Conference on Stochastic Processes and their Applications SPA 2025 (July 14-18, Wroclaw, Poland).
- Anne Gégout-Petit was chair of the program committee of ENBIS 2024 (Leuven, 15-19 september)
- U. Herbach is member of the scientific committee of the "Statistical Methods for Post Genomic Data" (SMPGD) annual congress.
- D. Villemonais co-organized the mini-symposium "Mathematical Insights into Telomere length dynamics" at ECMTB 2024, Toledo, Spain.

9.1.2 Member of the editorial boards

- N. Champagnat is associate editor for *ESAIM: Probability & Statistics* and *Stochastic Models*
- D. Villemonais is associate editor for *Applied Probability Trust*

9.1.3 Invited talks in conferences or workshops, seminar talks

- V. Brodu has been invited to give a talk at the MoVi seminar in Rennes in January, and online at the GDR MORPHEA together with C. Fritsch in April. He also gave a talk

- N. Champagnat has been invited to give talks at the Third International Biostochastic Workshop in Valparaíso, Chile in March and at the Journée de la Chaire MMB at the VEOLIA company headquarters in Paris in October.

- C. Fritsch has been invited to give a talk in a minisymposium at the European Conference on Mathematical and Theoretical Biology in Toledo in July and at the GDR MORPHEA in April.

- U. Herbach has been invited to give talks at the University of British Columbia Mathbio seminar (online) in January, at the GT Bioss annual workshop in Paris in May, at the SBDM team seminar in LBMC, Lyon in June, and at the INRAE MaIAGE seminar in Jouy-en-Josas in October.

- E. Strickler gave talks at the Séminaire de Probabilités de Rennes and at Séminaire de Probabilités de Toulouse, both in November.

- D. Villemonais has been invited to give a talk at IWBPA 2024 in Extremadura, Spain in April and at the 16ème Colloque Franco-Roumain in Bucharest, Romania in August.

- S. Wantz-Mézières gave a talk at the European Network for Business and Industrial Statistics Conference ENBIS-2024, in September.

9.1.4 Contributed talks, colloquium talks, posters

- N. Champagnat gave a colloquium talk at the Rhein-Main-Kolloquium Stochastik in Frankfurt, Germany in July.

- S. Ferrigno has presented a poster at the 18th International Joint Conference on Computational and Financial Econometrics (CFE) and Computational and Methodological Statistics, CFE-CMStatistics 2024 in London (King's College) in December.

- M. Gaillard gave talks at the "55ème Journées des Statistiques" in Bordeaux in May, at the "Congrès des Jeunes Chercheur.e.s en Mathématiques appliquées" in Lyon in October, and at the "24ème Forum des jeunes mathématiciennes et mathématiciens" in Montpellier in November.

- V. Brodu participated to a poster session at JMBS 2024 in Nantes in June, and also gave a talk at the MMEE 2024 Conference in Vienna in July and at the "Congrès des Jeunes Chercheur.e.s en Mathématiques appliquées" in Lyon in October

9.1.5 Scientific expertise

- N. Champagnat evaluated a research project submitted to Shape-Med@Lyon and the ISF.

- N. Champagnat has been a member of the Committee for junior permanent research positions of Centre Inria de Grenoble and Centre Inria de Bordeaux.

- N. Champagnat and C. Fritsch were members of the Committee for career advancement of Inria personnel.

- C. Fritsch has been a member of the Committee for junior permanent research positions of Centre Inria de Saclay and Centre Inria de Paris.

- C. Fritsch has been a member of the Committee for career advancement of Inria personnel.

9.1.6 Research administration

- V. Brodu is an elected representative of doctoral students at the doctoral school committee (local scale), and also at the doctoral college committee (regional scale).
- N. Champagnat is elected member of the Commission d'Evaluation of Inria since September, member of the COMIPERS (hiring committee for non-permanent positions) of Centre Inria de Nancy, substitute member of the Comité de Centre of Centre Inria de Nancy and local researcher (correspondant local) representing the COERLE (Inria's Ethic Committee) at Centre Inria de Nancy.
- C. Fritsch is an elected member of the Commission d'Evaluation of Inria.
- A. Gégout-Petit is director of the research unit IECL (Institut Elie Cartan de Lorraine), Mathematics laboratory of Univ. Lorraine (200 members).

9.2 Teaching - Supervision - Juries

9.2.1 Teaching

- S. Ferrigno is in charge of the "DU Big Data & Data Science" in ENSMN, Univ. Lorraine.
- D. Villemonais was the head of the Mathematical Engineering Major of ENSMN, Univ. Lorraine up to June 2024.
- Licence: S. Baland, Probabilities, 20h, L2 Informatique, Univ. Lorraine.
- Licence: S. Baland, Analysis, 12h, L2 CPU, Univ. Lorraine.
- Licence: S. Baland, Probability Theory, 40h, L3, first year of ENSMN, Univ. Lorraine.
- Licence: V. Brodu, Statistical inference, 40h, L3, first year of ENSMN, Univ. Lorraine.
- Licence: V. Brodu, Numerical Analysis, 20h, L3, first year of ENSMN, Univ. Lorraine.
- Master: N. Champagnat, Introduction to Quantitative Finance, 12h, M1, second year of ENSMN, Univ. Lorraine.
- Master: N. Champagnat, Introduction to Quantitative Finance, 9h, M2, third year of ENSMN, Univ. Lorraine.
- Master: S. Ferrigno, Experimental designs, 6h, M1, fourth year of EEIGM, Univ. Lorraine.
- Master: S. Ferrigno, Data analyzing and mining, 36h, M1, second year of ENSMN, Univ. Lorraine.
- Master: S. Ferrigno, Modeling and forecasting, 32h, M1, second year of ENSMN, Univ. Lorraine.
- Master: S. Ferrigno, Training projects, 18h, M1/M2, second and third year of ENSMN, Univ. Lorraine.
- Licence: S. Ferrigno, Descriptive and inferential statistics, 60h, L2, second year of EEIGM, Univ. Lorraine.
- Licence: S. Ferrigno, Statistical modeling, 60h, L2, second year of EEIGM, Univ. Lorraine.
- Licence: S. Ferrigno, Mathematical and computational tools, 20h, L3, third year of EEIGM, Univ. Lorraine.
- Licence: S. Ferrigno, Training projects, 40h, L1/L3, first, second and third year of EEIGM, Univ. Lorraine.
- Master: C. Fritsch, Inverse problem, 18h, M1, second year of ENSMN, Univ. Lorraine.
- Licence: C. Fritsch, Probability Theory, 28h, L3, first year of ENSMN, Univ. Lorraine.

- License: M. Gaillard, Numerical analysis and Optimization tutorial, 23h, L3, first year of ENSMN, Univ. Lorraine.
- License: M. Gaillard, Statistical inference, 21h, L3, first year of ENSMN, Univ. Lorraine.
- Master: A. Gégout-Petit, Statistics, modeling, data analysis, 80h, master in applied mathematics, Univ. Lorraine.
- Licence: V. Kagan, Probability theory tutorial, 40h, L3, first year of ENSMN, Univ. Lorraine.
- Licence: V. Kagan, Numerical Analysis tutorial, 20h, L3, first year of ENSMN, Univ. Lorraine.
- Licence: A. Rago, Mathematics, 35h, L3, first year of ENSMN, Univ. Lorraine.
- Master: A. Rago, Data analyzing and mining, 18h, M1, second year of ENSMN, Univ. Lorraine.
- Master: A. Rago, Introduction to machine learning, 12h, M1, second year of ENSMN, Univ. Lorraine.
- Master: A. Rago, Statistiques pour la grande dimension, 20h, M2 IMSD/third year of ENSMN, Univ. Lorraine.
- Master: E. Strickler, Processus Stochastiques Discrets, 18h, Master 2 MFA, Université de Lorraine.
- Master: V. Vidhi , Analyse de données, 18h, M1, second year of ENSMN, Univ. Lorraine.
- Master: D. Villemonais, Probability Theory, 28h, M1, Master of mathematics, Univ. Strasbourg.
- Master: S. Wantz-Mézières, Learning and analysis of medical data, 36h, with J.M. Moureaux, M2 SNIM, Univ. Lorraine.
- Licence: S. Wantz-Mézières, Applied mathematics for management, financial mathematics, Probability and Statistics, 160h, IUT Nancy-Charlemagne (L1/L2/L3), Univ. Lorraine.
- Licence: S. Wantz-Mézières, Probability, 100h, first year in TELECOM Nancy, Univ. Lorraine.

9.2.2 Supervision

PhD

- PhD in progress: Sophie Baland, “Telomere length dynamics : modelisation, estimation and application to diagnostic support systems”, funding LUE, since September 2023. Advisors: S. Toupance (Univ. Lorraine) and D. Villemonais.
- PhD in progress: Virgile Brodu,“Stochastic individual-based models with allometric dynamics: survival, convergence, biological implications.”, grant ENS Lyon, since September 2022. Advisors: S. Billiard (Univ. Lille), N. Champagnat, C. Fritsch.
- PhD in progress: Mathilde Gaillard, “Processus de Markov déterministes par morceaux et inférence bayésienne de réseaux de gènes”, grant PEPR Santé Numérique, since October 2023. Advisors: A. Gégout-Petit, U. Herbach.
- PhD in progress: Anouar Jeddi, “Convergence of individual-based population models to Hamilton-Jacobi equations”, grant ERC SINGER (Ecole Polytechnique), since September 2023. Advisors: S. Méléard (Ecole Polytechnique) and N. Champagnat.
- PhD in progress: Vincent Kagan, “Asymptotic behavior of epidemiological epidemiological models with individual viral load”, funding Université de Lorraine, since September 2023. Advisors: E. Strickler (Univ. Lorraine) and D. Villemonais.
- PhD in progress: Anouk Rago, “Inférence de réseaux de gènes dynamiques et prédiction d’expériences d’interventions biologiques dans des cellules cancéreuses”, grant Région Grand-Est and Inria, since October 2021. Advisors: N. Champagnat, A. Gégout-Petit.

- PhD in progress: Vidhi Vidhi, “Stochastic modeling and statistics for quantifying the evolution of tumor heterogeneity in chronic lymphocytic leukemia”, funding ITMO Cancer, since October 2024. Advisors: N. Champagnat, C. Fritsch, U. Herbach.

Other

- M2 Research project and Internship: Laetitia Bau, “Investigation de méthodes d’apprentissage statistique de type deep learning pour la création d’un outil d’assistance à la chirurgie éveillée des gliomes diffus de bas grade”, Advisor: S. Wantz-Mézières and J.M. Moureaux.
- M2 internship: Pascaline Kouda (M2 IMSD, Univ. Lorraine), “Évaluation et Amélioration des Modèles Diagnostiques de l’Insuffisance Cardiaque aux Urgences”. Advisors: Anne Gégout-Petit and Denis Villemonais.
- M2 internship: Vidhi Vidhi (Erasmus Mundus Double Masters Degree Programme, Masters of Science, Université Côte d’Azur), “Stochastic modeling and statistics for quantifying the evolution of tumor heterogeneity in chronic lymphocytic leukemia” funding ITMO Cancer. Advisors: N. Champagnat, C. Fritsch and U. Herbach.
- ENSMN third year (M2) Research project: Hamza Baroa, “Modèle statistique à variables latentes et inférence variationnelle pour l’analyse de données ordinales multivariées”, Advisor: U. Herbach
- M2 IMSD Research project: Thomas Flocquet and Cheikhadine Saleh Mahamat , "Analyse de données issues de cytométrie en flux", Advisor: A. Gégout-Petit and S. Wantz-Mézières
- M2 IMSD Research project: Yannis Raclot, Vicnesh Venedittan and Arthur Voranget, "Application de méthodes de réduction de dimension sur des données médicales", Advisor: A. Gégout-Petit and S. Wantz-Mézières
- ENSMN second year (M1) Research project: Edouard Ailloud-McIntyre, Louis Colomina, "Chaînes de Markov en temps continu et processus de Feller". Advisor: M. Gaillard
- ENSMN second year (M1) Research project: Antonin Clerc, “Emergence des allométries dans les écosystèmes”, Advisors: C. Fritsch and V. Brodu
- ENSMN second year (M1) Research project: Etienne Dedebant,"Détection de l’insuffisance rénale chronique chez le chat"(full-year research project). Advisors: D. Villemonais and S. Ferrigno.
- ENSMN second year (M1) Research project: Vincent Philisot,"Hybridization of Data Analysis and Genetic Algorithms to solve a Traveling Salesman Problem" (full-year research project). Advisors: W. Ramdane-Chérif and S. Ferrigno.

9.2.3 Juries

- N. Champagnat was referee for the PhD thesis of Lucas Journel (Sorbonne Univ., 03/06/2024) and for the HDR degree of Boris Nectoux (Univ. Clermont-Auvergne, 22/11/2024). He was also examiner for the HDR degree of Hélène Leman (Univ. Lyon, 07/06/2024).
- M. Gaillard was in the regional jury of the "Tournoi des Jeunes Mathematiciens et Mathematiciennes" in Nancy in April.
- E. Strickler was member of the jury for the oral of the Agrégation de Mathématiques in Strasbourg in June.
- D. Villemonais was an examiner of the jury for the PhD thesis of Elie Cerf (University Paris XIII - Sorbonne Paris-Nord, 13/12/2024) and for the PhD thesis of Brieuc Frénais (University of Strasbourg, 20/12/2024).

9.3 Popularization

J.-M. Monnez wrote lecture notes [28] on the interpretation of canonical analysis of two random vectors as a projected PCA, an extended Oja process for estimating eigenvectors and stochastic approximation for streaming canonical correlation, factorial correspondence and factorial discriminant analyses.

9.3.1 Specific official responsibilities in science outreach structures

- S. Ferrigno: Advisor of three groups of EEIGM students in the context of "La main à la Pâte" projects and "CGénial" projects, at Collèges Paul Verlaine in Malzéville, at Lycée La Craffe in Nancy, at Institut médico-éducatif (IME) in Commercy and in elementary schools in Nancy.
- S. Ferrigno: Advisor of a group of EEIGM students, "Ateliers expérimentaux : Mécanique et Statistique" Project, various high schools, Nancy.

9.3.2 Participation in Live events

- U. Herbach and N. Seyler gave a joint presentation of their ADT project Harissa at the "Café'In" general public seminar of Inria Nancy in October.

9.3.3 Others science outreach relevant activities

- C. Fritsch gave two talks as part of the "Chiche!" program at Lycée Jean XXIII, Metz, in November.
- V. Brodu supervised a maths club for highschool students in Lycée Jeanne d'Arc, Nancy. This club hosted a dozen students for two hours sessions, on a weekly basis. We created a team of six active highschool students, and participated to the national mathematical tournament TFJM. We won the tournament, and went to the european edition ETEAM, where we were ranked second.

10 Scientific production

10.1 Major publications

[1] R. Azaïs, S. Ferrigno and M.-J. Martinez. 'cvmgof: an R package for Cramér-von Mises goodness-of-fit tests in regression models'. In: *Journal of Statistical Computation and Simulation* 92.6 (2022), pp. 1246–1266. DOI: 10.1080/00949655.2021.1991346. URL: https://hal.science/hal-03101612 (cit. on p. 7).

[2] N. Champagnat and V. Hass. 'Convergence of population processes with small and frequent mutations to the canonical equation of adaptive dynamics'. In: *The Annals of Applied Probability* 35.1 (1st Feb. 2025), pp. 1–63. DOI: 10.1214/24-AAP2103. URL: https://hal.science/hal-04034027 (cit. on pp. 5, 8).

[3] N. Champagnat, S. Méléard and V. C. Tran. 'Multiscale eco-evolutionary models: from individuals to populations'. In: International Congress of Mathematicians, ICM 2022. Vol. 7. 1. fully virtually, Russia: EMS Press, 15th Dec. 2023, pp. 5656–5678. DOI: 10.4171/ICM2022/24. URL: https://inria.hal.science/hal-04407958.

[4] N. Champagnat and D. Villemonais. 'General criteria for the study of quasi-stationarity'. In: *Electronic Journal of Probability* (2023). DOI: 10.1214/22-EJP880. URL: https://inria.hal.science/hal-01672979. In press (cit. on pp. 9, 14).

[5] E. Eyraud, E. Maurat, J.-M. Sac-Epee, P. Henrot, M. Zysman, P. Esteves, T. Trian, H. Bégueret, P.-O. Girodet, M. Thumerel, R. Hustache-Castaing, R. Marthan, F. Levet, P. Vallois, C. Contin-Bordes, P. Berger and I. Dupin. 'Short-range interactions between fibrocytes and CD8+ T cells in COPD bronchial inflammatory response'. In: *eLife* (21st Oct. 2022). DOI: 10.1101/2022.10.21.513138. URL: https://hal.science/hal-03958836.

[6] C. Fritsch, M. Grosdidier, A. Gégout-Petit and B. Marçais. *Mechanistic-statistical model for the expansion of ash dieback*. 6th Sept. 2024. URL: https://hal.science/hal-04690647.

[7] C. Fritsch, D. Villemonais and N. Zalduendo. 'The Multi-type Bisexual Galton-Watson Branching Process'. In: *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* (2024), 36p. DOI: 10.48550/arXiv.2206.09622. URL: https://hal.science/hal-03696115. In press.

[8] J.-M. Monnez. 'Stochastic approximation of eigenvectors and eigenvalues of the Q-symmetric expectation of a random matrix'. In: *Communications in Statistics - Theory and Methods* 53.5 (2024), pp. 1669–1683. DOI: 10.1080/03610926.2022.2107225. URL: https://hal.science/hal-03956687 (cit. on p. 20).

[9] S. Toupance, D. Villemonais, D. Germain, A. Gégout-Petit, E. Albuisson and A. Benetos. 'The individual's signature of telomere length distribution'. In: *Scientific Reports* 9.1 (24th Jan. 2019), pp. 1–8. DOI: 10.1038/s41598-018-36756-8. URL: https://inria.hal.science/hal-01925000 (cit. on p. 10).

[10] E. Ventre, U. Herbach, T. Espinasse, G. Benoit and O. Gandrillon. 'One model fits all: Combining inference and simulation of gene regulatory networks'. In: *PLoS Computational Biology* 19.3 (27th Mar. 2023), e1010962. DOI: 10.1371/journal.pcbi.1010962. URL: https://inria.hal.science/hal-04071033 (cit. on p. 7).

10.2 Publications of the year

International journals

[11] C. Fritsch, D. Villemonais and N. Zalduendo. 'The Multi-type Bisexual Galton-Watson Branching Process'. In: *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* 60.4 (2024), pp. 2975–3008. DOI: 10.1214/23-AIHP1417. URL: https://hal.science/hal-03696115 (cit. on p. 14).

[12] A. Koshkin, U. Herbach, M. R. Martínez, O. Gandrillon and F. Crauste. 'Stochastic modeling of a gene regulatory network driving B cell development in germinal centers'. In: *PLoS ONE* 19.3 (28th Mar. 2024), e0301022. DOI: 10.1371/journal.pone.0301022. URL: https://hal.science/hal-04490302 (cit. on pp. 11, 18).

[13] R. Loubaton, N. Champagnat, P. Vallois and L. Vallat. 'MultiRNAflow: integrated analysis of temporal RNA-seq data with multiple biological conditions'. In: *Bioinformatics* 40.5 (29th May 2024), p. 4. DOI: 10.1093/bioinformatics/btae315. URL: https://inria.hal.science/hal-04407596 (cit. on p. 17).

[14] A. J. Martin, A.-M. Revol-Junelles, J. Petit, C. Gaiani, M. Leyva Salas, N. Nourdin, M. Khatbane, P. Mafra de Almeida Costa, S. Ferrigno, B. Ebel, M. Schivi, A. Elfassy, C. Mangavel and F. Borges. 'Deciphering rind color heterogeneity of smear-ripened Munster cheese and its Association with microbiota'. In: *Foods* 13.14 (16th July 2024), p. 2233. DOI: 10.3390/foods13142233. URL: https://hal.univ-lorraine.fr/hal-04691925 (cit. on p. 19).

[15] J.-M. Monnez. 'Stochastic approximation of eigenvectors and eigenvalues of the Q-symmetric expectation of a random matrix'. In: *Communications in Statistics - Theory and Methods* 53.5 (2024), pp. 1669–1683. DOI: 10.1080/03610926.2022.2107225. URL: https://hal.science/hal-03956687.

[16] A. Prodhomme and E. Strickler. 'Large population asymptotics for a multitype stochastic SIS epidemic model in randomly switching environment'. In: *The Annals of Applied Probability* 34.3 (1st June 2024), pp. 3125–3180. DOI: 10.1214/23-AAP2035. URL: https://hal.science/hal-04780354 (cit. on p. 15).

Reports & preprints

[17] M. Benaim, C. Lobry, T. Sari and E. Strickler. *Dispersal-induced growth or decay in a time-periodic environment*. 27th Mar. 2024. URL: https://hal.inrae.fr/hal-04523334 (cit. on p. 15).

[18] M. Benaïm, C. Lobry, T. Sari and E. Strickler. *Population growth on a time varying network*. 12th Nov. 2024. URL: https://hal.science/hal-04779418 (cit. on p. 15).

[19] A. Benetos, C. Fritsch, E. Horton, L. Lenotre, S. Toupance and D. Villemonais. *Stochastic branching models for the telomeres dynamics in a model including telomerase activity*. 15th July 2024. URL: https://hal.science/hal-04648211 (cit. on p. 17).

[20] S. Billiard, V. Brodu, N. Champagnat and C. Fritsch. *An individual-based stochastic model reveals strong constraints on allometric relationships with minimal metabolic and ecological assumptions*. 21st Jan. 2025. URL: https://hal.science/hal-04904545 (cit. on p. 17).

[21] N. Champagnat, A. Gégout-Petit and A. Rago. *Semi-Lasso: a weighted Lasso designed for the integration of known regressors in linear model*. Mar. 2024. URL: https://inria.hal.science/hal-04517255 (cit. on p. 17).

[22] N. Champagnat, T. Lelièvre, M. Ramil, J. Reygner and D. Villemonais. *Quasi-stationary distribution for kinetic SDEs with low regularity coefficients*. 1st Oct. 2024. URL: https://hal.science/hal-04720688 (cit. on p. 14).

[23] M. Costa and E. Strickler. *Corrigendum to : "A piecewise deterministic model for a prey-predator community" An extinction-persistence issue*. 16th Dec. 2024. URL: https://hal.science/hal-04839491.

[24] C. Fritsch, M. Grosdidier, A. Gégout-Petit and B. Marçais. *Mechanistic-statistical model for the expansion of ash dieback*. 6th Sept. 2024. URL: https://hal.science/hal-04690647 (cit. on pp. 11, 21).

[25] C. Fritsch, D. Villemonais and N. Zalduendo. *Quasi-limiting behaviour of the sub-critical Multitype Bisexual Galton-Watson Branching Process*. 27th June 2024. URL: https://hal.science/hal-04628438 (cit. on p. 14).

[26] P. Monmarché, S. J. Schreiber and É. Strickler. *Impacts of Tempo and Mode of Environmental Fluctuations on Population Growth: Slow- and Fast-Limit Approximations of Lyapunov Exponents for Periodic and Random Environments*. 22nd Aug. 2024. URL: https://hal.science/hal-04769603 (cit. on p. 15).

Other scientific publications

[27] S. Ferrigno. 'Nonparametric estimation of reference curves'. In: CMStatistics 2024. Londres, United Kingdom, 14th Dec. 2024. URL: https://hal.science/hal-04832287 (cit. on p. 20).

Educational activities

[28] J.-M. Monnez. 'Streaming data analysis Canonical analysis of two random vectors'. Doctoral. France, 21st Jan. 2025. URL: https://hal.science/hal-04908126 (cit. on p. 30).

Software

[29] [SW] U. Herbach, *Harissa: tools for mechanistic gene network inference from single-cell data* version 3.0.12, 2024. LIC: BSD 3-Clause. HAL: ⟨hal-03370296⟩, URL: https://hal.science/hal-03370296, VCS: https://github.com/ulysseherbach/harissa, SWHID: ⟨swh:1:dir:8751b2c933af3332d28e2f02d8338ca4e6fdcf32;origin=https://github.com/ulysseherbach/harissa;visit=swh:1:snp:0e96961dfcdcaac02e5769a28065318db3407ae7;anchor=swh:1:rev:9b2e208bcc3183237423b8c8ebe347fa32e4ac7d⟩.

10.3 Cited publications

[30] D. F. Anderson, A. Ganguly and T. G. Kurtz. 'Error analysis of tau-leap simulation methods'. In: *Ann. Appl. Probab.* 21.6 (2011), pp. 2226–2262. DOI: 10.1214/10-AAP756. URL: https://doi.org/10.1214/10-AAP756 (cit. on p. 9).

[31] R. Azaïs, S. Ferrigno and M.-J. Martinez. 'cvmgof: Cramer-von Mises goodness-of-fit tests'. An R-package, available on the CRAN. Nov. 2018. URL: https://hal.archives-ouvertes.fr/hal-02014516 (cit. on p. 6).

[32] B. Bastien, T. Boukhobza, H. Dumond, A. Gégout-Petit, A. Muller-Gueudin and C. Thiébaut. 'A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology'. https://arxiv.org/abs/1909.05481 - working paper or preprint. Sept. 2019. URL: https://hal.archives-ouvertes.fr/hal-02173568 (cit. on p. 6).

[33] M. Benaim, B. Cloez and F. Panloup. 'Stochastic approximation of quasi-stationary distributions on compact spaces and applications'. In: *Ann. Appl. Probab.* 28.4 (2018), pp. 2370–2416. DOI: 10.1214/17-AAP1360. URL: https://doi.org/10.1214/17-AAP1360 (cit. on p. 9).

[34] M. Benaim, C. Lobry, T. Sari and E. Strickler. 'When can a population spreading across sink habitats persist?' In: *Journal of Mathematical Biology* 88.2 (2024), p. 19 (cit. on p. 15).

[35] M. Benaïm, N. Champagnat and D. Villemonais. 'Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain'. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 57.2 (2021), pp. 726–739. DOI: 10.1214/20-AIHP1093. URL: https://doi.org/10.1214/20-AIHP1093 (cit. on p. 9).

[36] M. Benaïm, C. Lobry, T. Sari and É. Strickler. 'Untangling the role of temporal and spatial variations in persistence of populations'. In: *Theoretical Population Biology* 154 (2023), pp. 1–26 (cit. on p. 15).

[37] A. Benetos et al. 'Short leukocyte telomere length precedes clinical expression of atherosclerosis: The blood-and-muscle model'. In: *Circulation research* 122.4 (Feb. 2018), pp. 616–623. DOI: 10.1161/CIRCRESAHA.117.311751 (cit. on p. 10).

[38] A. Bonnaffoux, U. Herbach, A. Richard, A. Guillemin, S. Gonin-Giraud, P.-A. Gros and O. Gandrillon. 'WASABI: a dynamic iterative framework for gene regulatory network inference'. In: *BMC Bioinformatics* 20.1 (2019), pp. 1–19 (cit. on p. 7).

[39] F. Campillo and N. Champagnat. 'Simulation and analysis of an individual-based model for graph-structured plant dynamics'. In: *Ecological Modelling* 234 (2012), pp. 93–105 (cit. on pp. 5, 6).

[40] F. Campillo and C. Fritsch. 'Weak convergence of a mass-structured individual-based model'. In: *Appl. Math. Optim.* 72.1 (2015), pp. 37–73. DOI: 10.1007/s00245-014-9271-3. URL: https://doi.org/10.1007/s00245-014-9271-3 (cit. on p. 6).

[41] H. Cardot, P. Cénac and J.-M. Monnez. 'A fast and recursive algorithm for clustering large datasets with k-medians'. In: *Computational Statistics & Data Analysis* 56.6 (2012), pp. 1434–1449 (cit. on pp. 7, 20).

[42] N. Champagnat. 'A microscopic interpretation for adaptive dynamics trait substitution sequence models'. In: *Stoch. Process. Appl.* 116.8 (2006), pp. 1127–1160 (cit. on pp. 5, 8).

[43] N. Champagnat, R. Ferrière and S. Méléard. 'From individual stochastic processes to macroscopic models in adaptive evolution'. In: *Stoch. Models* 24.suppl. 1 (2008), pp. 2–44. DOI: 10.1080/15326340802437710. URL: http://dx.doi.org/10.1080/15326340802437710 (cit. on p. 9).

[44] N. Champagnat, R. Ferrière and S. Méléard. 'Unifying evolutionary dynamics: From individual stochastic processes to macroscopic evolution'. In: *Theor. Pop. Biol.* 69 (2006), pp. 297–321 (cit. on p. 6).

[45] N. Champagnat and P.-E. Jabin. 'The evolutionary limit for models of populations interacting competitively via several resources'. In: *J. Differential Equations* 251.1 (2011), pp. 176–195. DOI: 10.1016/j.jde.2011.03.007. URL: https://doi.org/10.1016/j.jde.2011.03.007 (cit. on p. 5).

[46] N. Champagnat and S. Méléard. 'Polymorphic evolution sequence and evolutionary branching'. In: *Probab. Theory Related Fields* 151.1-2 (2011), pp. 45–94. DOI: 10.1007/s00440-010-0292-9. URL: https://doi.org/10.1007/s00440-010-0292-9 (cit. on pp. 5, 8).

[47] N. Champagnat, S. Méléard and V. C. Tran. 'Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer'. In: *The Annals of Applied Probability* 31.4 (2021), pp. 1820–1867 (cit. on p. 5).

[48] N. Champagnat and D. Villemonais. 'Convergence of the Fleming-Viot process toward the minimal quasi-stationary distribution'. In: *ALEA - Latin American Journal of Probability and Mathematical Statistics* (2019). to appear (cit. on p. 9).

[49] N. Champagnat and D. Villemonais. 'Exponential convergence to quasi-stationary distribution and Q-process'. In: *Probab. Theory Related Fields* 164.1-2 (2016), pp. 243–283. DOI: 10.1007/s00440-014-0611-7. URL: https://doi.org/10.1007/s00440-014-0611-7 (cit. on p. 9).

[50] B. Cloez and C. Fritsch. 'Gaussian approximations for chemostat models in finite and infinite dimensions'. In: *J. Math. Biol.* 75.4 (2017), pp. 805–843. DOI: 10.1007/s00285-017-1097-6. URL: https://doi.org/10.1007/s00285-017-1097-6 (cit. on p. 9).

[51] D. A. Dawson. 'Measure-valued Markov processes'. In: *École d'Été de Probabilités de Saint-Flour XXI—1991*. Vol. 1541. Lecture Notes in Math. Berlin: Springer, 1993, pp. 1–260 (cit. on p. 8).

[52] U. Dieckmann and R. Law. 'The dynamical theory of coevolution: a derivation from stochastic ecological processes'. In: *J. Math. Biol.* 34.5-6 (1996), pp. 579–612 (cit. on p. 8).

[53] O. Diekmann, P.-E. Jabin, S. Mischler and B. Perthame. 'The dynamics of adaptation: An illuminating example and a Hamilton-Jacobi approach'. In: *Theor. Pop. Biol.* 67 (2005), pp. 257–271 (cit. on p. 8).

[54] W. Ding and T. Giletti. 'Admissible speeds in spatially periodic bistable reaction-diffusion equations'. In: *Advances in Mathematics* 389 (2021), p. 107889 (cit. on p. 5).

[55] K. Duarte, J.-M. Monnez and E. Albuisson. 'Sequential linear regression with online standardized data'. In: *Plos one* 13.1 (2018), e0191186 (cit. on p. 20).

[56] G. R. Ducharme and S. Ferrigno. 'An omnibus test of goodness-of-fit for conditional distributions with applications to regression models'. In: *J. Statist. Plann. Inference* 142.10 (2012), pp. 2748–2761. DOI: 10.1016/j.jspi.2012.04.008. URL: https://doi.org/10.1016/j.jspi.2012.04.008 (cit. on p. 7).

[57] S. N. Ethier and T. G. Kurtz. 'Fleming-Viot processes in population genetics'. In: *SIAM J. Control Optim.* 31.2 (1993), pp. 345–386 (cit. on p. 8).

[58] C. Fritsch, J. Harmand and F. Campillo. 'A modeling approach of the chemostat'. In: *Ecological Modelling* (2014) (cit. on p. 9).

[59] A. Gégout-Petit, A. Gueudin-Muller and C. Karmann. 'The revisited knockoffs method for variable selection in L 1-penalized regressions'. In: *Communications in Statistics - Simulation and Computation* 0.0 (2020), pp. 1–14. DOI: 10.1080/03610918.2020.1775850. eprint: https://doi.org/10.1080/03610918.2020.1775850. URL: https://doi.org/10.1080/03610918.2020.1775850 (cit. on p. 6).

[60] U. Herbach, A. Bonnaffoux, T. Espinasse and O. Gandrillon. 'Inferring gene regulatory networks from single-cell data: a mechanistic approach'. In: *BMC Syst. Biol.* 11.1 (2017) (cit. on pp. 7, 18).

[61] P.-E. Jabin. 'Small populations corrections for selection-mutation models'. In: *Netw. Heterog. Media* 7.4 (2012), pp. 805–836. DOI: 10.3934/nhm.2012.7.805. URL: https://doi.org/10.3934/nhm.2012.7.805 (cit. on p. 9).

[62] G. Katriel. 'Dispersal-induced growth in a time-periodic environment'. In: *Journal of Mathematical Biology* 85.3 (2022), p. 24 (cit. on pp. 14, 15).

[63] M. Klann, A. Ganguly and H. Koeppl. 'Hybrid spatial Gillespie and particle tracking simulation'. In: *Bioinformatics* 28.18 (2012), pp. i549–i555 (cit. on p. 9).

[64] B. Lalloué, J.-M. Monnez and E. Albuisson. 'Streaming constrained binary logistic regression with online standardized data'. In: *Journal of Applied Statistics* (2021). In press. DOI: 10.1080/02664763.2020.1870672 (cit. on p. 7).

[65] S. Lee, M. Ramil and I. Seo. 'Asymptotic stability and cut-off phenomenon for the underdamped Langevin dynamics'. In: *arXiv preprint arXiv:2311.18263* (2023) (cit. on p. 14).

[66] A. Lorz, S. Mirrahimi and B. Perthame. 'Dirac mass dynamics in multidimensional nonlocal parabolic equations'. In: *Comm. Partial Differential Equations* 36.6 (2011), pp. 1071–1098. DOI: 10.1080/03605302.2010.538784. URL: http://dx.doi.org/10.1080/03605302.2010.538784 (cit. on p. 8).

[67] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. J. A. Jacobs and J. S. van Heerwaarden. 'Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction'. In: *Stochastic and spatial structures of dynamical systems (Amsterdam, 1995)*. Konink. Nederl. Akad. Wetensch. Verh. Afd. Natuurk. Eerste Reeks, 45. Amsterdam: North-Holland, 1996, pp. 183–231 (cit. on p. 8).

[68] S. Mirrahimi, G. Barles, B. Perthame and P. E. Souganidis. 'A singular Hamilton-Jacobi equation modeling the tail problem'. In: *SIAM J. Math. Anal.* 44.6 (2012), pp. 4297–4319. DOI: 10.1137/100819527. URL: https://doi.org/10.1137/100819527 (cit. on p. 9).

[69] P. Monmarché and E. Strickler. 'Asymptotic expansion of the invariant measurefor Markov-modulated ODEs at high frequency'. In: *arXiv e-prints*, arXiv:2309.16464 (Sept. 2023), arXiv:2309.16464. DOI: 10.48550/arXiv.2309.16464. arXiv: 2309.16464 [math.PR] (cit. on p. 15).

[70] J.-M. Monnez. 'Approximation stochastique en analyse factorielle multiple'. In: *Ann. I.S.U.P.* 50.3 (2006), pp. 27–45 (cit. on p. 7).

[71] J.-M. Monnez. 'Convergence d'un processus d'approximation stochastique en analyse factorielle'. In: *Publ. Inst. Statist. Univ. Paris* 38.1 (1994), pp. 37–55 (cit. on p. 7).

[72] J.-M. Monnez. 'Stochastic approximation of the factors of a generalized canonical correlation analysis'. In: *Statist. Probab. Lett.* 78.14 (2008), pp. 2210–2216. DOI: 10.1016/j.spl.2008.01.088. URL: http://dx.doi.org/10.1016/j.spl.2008.01.088 (cit. on p. 7).

[73] J.-M. Monnez and A. Skiredj. 'Convergence of a normed eigenvector stochastic approximation process and application to online principal component analysis of a data stream'. working paper or preprint. May 2019. URL: https://hal.archives-ouvertes.fr/hal-01844419 (cit. on p. 7).

[74] B. Perthame and G. Barles. 'Dirac concentrations in Lotka-Volterra parabolic PDEs'. In: *Indiana Univ. Math. J.* 57.7 (2008), pp. 3275–3301. DOI: 10.1512/iumj.2008.57.3398. URL: http://dx.doi.org/10.1512/iumj.2008.57.3398 (cit. on p. 8).

[75] B. Perthame and M. Gauduchon. 'Survival thresholds and mortality rates in adaptive dynamics: conciliating deterministic and stochastic simulations'. In: *Math. Med. Biol.* 27.3 (2010), pp. 195–210. DOI: 10.1093/imammb/dqp018. URL: https://doi.org/10.1093/imammb/dqp018 (cit. on p. 9).

[76] P. C. de Raynal, S. Menozzi, A. Pesce and X. Zhang. 'Heat kernel and gradient estimates for kinetic SDEs with low regularity coefficients'. In: *Bulletin des Sciences Mathématiques* 183 (2023), p. 103229 (cit. on p. 14).

[77] A. Richard, L. Boullu, U. Herbach, A. Bonnafoux, V. Morin, E. Vallin, A. Guillemin, N. Papili Gao, R. Gunawan, J. Cosette, O. Arnaud, J.-J. Kupiec, T. Espinasse, S. Gonin-Giraud and O. Gandrillon. 'Single-cell-based analysis highlights a surge in cell-to-cell molecular variability preceding irreversible commitment in a differentiation process'. In: *PLOS Biology* 14.12 (2016) (cit. on p. 11).

[78] D. Villemonais. 'Interacting particle systems and Yaglom limit approximation of diffusions with unbounded drift'. In: *Electron. J. Probab.* 16 (2011), no. 61, 1663–1692. DOI: 10.1214/EJP.v16-925. URL: https://doi.org/10.1214/EJP.v16-925 (cit. on p. 9).

[79] D. Waxman and S. Gavrilets. '20 questions on adaptive dynamics'. In: *J. Evol. Biol.* 18 (2005), pp. 1139–1154 (cit. on p. 9).