RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2019

Section Application Domains

Edition: 2020-03-21

COMPUTATIONAL BIOLOGY
1. ABS Project-Team (section vide) . . . 5
2. BEAGLE Project-Team . . . 6
3. BIGS Project-Team . . . 7
4. CAPSID Project-Team . . . 8
5. DYLISS Project-Team . . . 10
6. ERABLE Project-Team . . . 11
7. GENSCALE Project-Team . . . 12
8. IBIS Project-Team (section vide) . . . 13
9. LIFEWARE Project-Team . . . 14
10. MORPHEME Project-Team (section vide) . . . 16
11. MOSAIC Project-Team (section vide) . . . 17
12. PLEIADE Project-Team . . . 18
13. SERPICO Project-Team . . . 22
COMPUTATIONAL NEUROSCIENCE AND MEDICINE
14. ARAMIS Project-Team . . . 24
15. ATHENA Project-Team . . . 25
16. BIOVISION Project-Team . . . 26
17. CAMIN Project-Team (section vide) . . . 27
18. EMPENN Project-Team . . . 28
19. EPIONE Project-Team (section vide) . . . 31
20. MATHNEURO Project-Team (section vide) . . . 32
21. MIMESIS Team . . . 33
22. MNEMOSYNE Project-Team . . . 34
23. NEUROSYS Project-Team . . . 35
24. OPIS Project-Team . . . 36
25. PARIETAL Project-Team . . . 39
EARTH, ENVIRONMENTAL AND ENERGY SCIENCES
26. AIRSEA Project-Team . . . 41
27. ANGE Project-Team . . . 42
28. CASTOR Project-Team (section vide) . . . 45
29. COFFEE Project-Team . . . 46
30. FLUMINANCE Project-Team (section vide) . . . 48
31. LEMON Project-Team (section vide) . . . 49
32. MAGIQUE-3D Project-Team . . . 50
33. SERENA Project-Team . . . 52
34. STEEP Project-Team . . . 53
35. TONUS Project-Team . . . 57
MODELING AND CONTROL FOR LIFE SCIENCES
36. BIOCORE Project-Team . . . 58
37. CARMEN Project-Team . . . 60

38. COMMEDIA Project-Team 61
39. DRACULA Project-Team (section vide) 62
40. M3DISIM Project-Team (section vide) 63
41. MAMBA Project-Team 64
42. MONC Project-Team 73
43. NUMED Project-Team (section vide) 74
44. REO Team 75
45. SISTM Project-Team 77
46. XPOP Project-Team 79

ABS Project-Team (section vide)

BEAGLE Project-Team

4. Application Domains

4.1. Functional and Evolutionary Biology

We do not distinguish our research and its application domains. Our shared idea is that the research is oriented by a scientific question, which is a multidisciplinary one, most often of biological nature. We do not develop methodologies independently from this question and then look for applications. Instead we collectively work with other disciplines to solve a question, with our competencies.

In consequence the application domains are already listed in the description of our projects and goals. They concern functional and evolutionary biology, related to critical social questions as human and plant health.

4.2. Social and Environmental Responsibility (Implication domains)

These last years we have maintained a frequent team discussion on the social and environmental responsibility of researchers. It has become more frequent this year, with the announcements of serious environmental issues by many governmental or non governmental organizations.

We are engaged in many actions regarding this responsibility. A constant ethics questioning, directing our research projects according to our values, teaching and popularization of ethical values. In particular we are engaged in several research projects on health and environment, and one of us has been a member of the institutional workgroup on environmental issues at Inria.

Regarding the functioning of research activities, we attempted a measure of the environmental footprint of our activities, regardless of their aims. It has the shape of a carbon footprint analysis, gathering the carbon footprint of travels, computer usage, computer equipment. We are aware of the incompleteness of this analysis, as well by not including many activities (nutrition, homeplace-workplace trips), and not taking other environmental issues than carbon emissions.

However it is a starting point, that we presented to many colleagues who were interested in reproducing the computation, so we give the headlines here. We used the unitary costs given by a website we constructed: https://ferme.yeswiki.net/Empreinte. The data was collected on an average taken over 3 years, 2016-2018 (we cannot yet at this stage make the analysis for 2019). Travels by members of the team or invited researchers emitted 39.86tCO2. Computing hours on local clusters emitted 17.46tCO2. The acquisition of computers accounts for the emission of 4.53tCO2. The total is 61.85tCO2.

Based on this lower bound of our CO2 emissions, the total per person is around 3 tons per person. Probably this number is highly underevaluated since it accounts for only part of our professional activities.

Whether it is too high or acceptable, and the establishment of a carbon budget is a difficult question. If we refer to the goals of the COP21 conclusions, we should emit less than 2 tons of CO2 per person in 2050 to reach carbon neutrality. This includes professional and personal life, and all the services we benefit from. We did not arrive yet at a consensus on the objective we should reach to consider we have a sutainable activity, but we in majority recognize that we are anyway far above what would be a consensus objective. We are still in a discussion to engage in a reduction.

BIGS Project-Team

4. Application Domains

4.1. Tumor growth-oncology

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

4.2. Genomic data and micro-organisms population

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

4.3. Epidemiology and e-health

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

4.4. Dynamics of telomeres

Telomeres are disposable buffers at the ends of chromosomes which are truncated during cell division; so that, over time, due to each cell division, the telomere ends become shorter. By this way, they are markers of aging. Trough a collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently obtained data on the distribution of the length of telomeres from blood cells. With members of Inria team TOSCA, we want to work in three connected directions: (1) refine methodology for the analysis of the available data; (2) propose a dynamical model for the lengths of telomeres and study its mathematical properties (long term behavior, quasi-stationarity, etc.); and (3) use these properties to develop new statistical methods. A slot of postdoc position is already planned in the Lorraine Université d'Excellence, LUE project GEENAGE (managed by CHU Nancy).

CAPSID Project-Team

4. Application Domains

4.1. Biomedical Knowledge Discovery

Participants: Marie-Dominique Devignes [contact person], Malika Smaïl-Tabbone [contact person], Sabeur Aridhi, David Ritchie, Gabin Personeni, Seyed Ziaeddin Alborzi, Kevin Dalleau, Bishnu Sarker, Emmanuel Bresso, Claire Lacomblez, Floriane Odje, Athénaïs Vaginay.

Our main application for Axis 1 : "New Approaches for Knowledge Discovery in Structural Databases", concerns biomedical knowledge discovery. We intend to develop KDD approaches on preclinical (experimental) or clinical datasets integrated with knowledge graphs with a focus on discovering which PPIs or molecular machines play an essentiel role in the onset of a disease and/or for personalized medicine.

As a first step we have been involved since 2015 in the ANR RHU “FIGHT-HF” (Fight Heart Failure) project, which is coordinated by the CIC-P (Centre d’Investigation Clinique Plurithématique) at the CHRU Nancy and INSERM U1116. In this project, the molecular mechanisms that underly heart failure (HF) are re-visited at the cellular and tissue levels in order to adapt treatments to patients’ needs in a more personalized way. The Capsid team is in charge of a workpackage dedicated to network science. A platform has been constructed with the help of a company called Edgeleap (Utrecht, NL) in which biological molecular data and ontologies, available from public sources, are represented in a single integrated complex network also known as knowledge graph. We are developing querying and analysis facilities to help biologists and clinicians interpreting their cohort results in the light of existing interactions and knowledge. We are also currently analyzing pre-clinical data produced at the INSERM unit on the comparison of aging process in obese versus lean rats. Using our expertise in receptor-ligand docking, we are investigating possible cross-talks between mineralocorticoid and other nuclear receptors.

Another application is carried out in the context of a UL-funded interdisciplinary project in collaboration with the CRAN laboratory. It concerns the study of the role of estrogen receptors in the development of gliobastoma tumors. The available data is high-dimensional but involves rather small numbers of samples. The challenge is to identify relevant sets of genes which are differentially expressed in various phenoptyped groups (w.r.t. gender, age, tumor grade). The objectives are to infer pathways involving these genes and to propose candidate models of tumor development which will be experimentally tested thanks to an ex-vivo experimental system available at the CRAN.

Finally, simulating biological networks will be important to understand biological systems and test new hypotheses. One major challenge is the identification of perturbations responsible for the transformation of a healthy system to a pathological one and the discovery of therapeutic targets to reverse this transformation. Control theory, which consists in finding interventions on a system in order to prevent it to go in undesirable states or to force it to converge towards a desired state, is of great interest for this challenge. It can be formulated as “How to force a broken system (pathological) to act as it should do (normal state)?”. Many formalisms are used to model biological processes, such as Differential Equations (DE), Boolean Networks (BN), cellular automata. In her PhD thesis, Athenaïs Vaginay investigates ways to find a BN fitting both the knowledge about topology and state transitions “inferred“ from experimental data. This step is known as “boolean function synthesis”. Our aim is to design automated methods for building biological networks and define operators to intervene on them[29]. Our approaches will be driven by knowledge and keep close connection with experimental data.

4.2. Prokaryotic Type IV Secretion Systems

Participants: Marie-Dominique Devignes [contact person], Isaure Chauvot de Beauchêne [contact person], Bernard Maigret, David Ritchie, Philippe Noel, Antoine Moniot, Dominique Mias-Lucquin.

Concerning Axis 2 : "Integrative Multi-Component Assembly and Modeling", our first application domain is related to prokaryotic type IV secretion systems.

Prokaryotic type IV secretion systems constitute a fascinating example of a family of nanomachines capable of translocating DNA and protein molecules through the cell membrane from one cell to another [36]. The complete system involves at least 12 proteins. The structure of the core channel involving three of these proteins has recently been determined by cryo-EM experiments for Gram-negative bacteria [47], [63]. However, the detailed nature of the interactions between the other components and the core channel remains to be found. Therefore, these secretion systems represent a family of complex biological systems that call for integrated modeling approaches to fully understand their machinery.
In the framework of the Lorraine Université d'Excellence (LUE-FEDER) "CITRAM" project we are pursuing our collaboration with Nathalie Leblond of the Genome Dynamics and Microbial Adaptation (DynAMic) laboratory (UMR 1128, Université de Lorraine, INRA) on the mechanism of horizontal transfer by integrative conjugative elements (ICEs) and integrative mobilisable elements (IMEs) in prokaryotic genomes. These elements use Type IV secretion systems for transferring DNA horizontally from one cell to another. We have discovered more than 200 new ICEs/IMEs by systematic exploration of 72 Streptococcus genomes and characterized a new class of relaxases [21]. We have modeled the dimer of this relaxase protein by homology with a known structure. For this, we have created a new pipeline to model symmetrical dimers of multi-domains proteins. As one activity of the relaxase is to cut the DNA for its transfer, we are also currently studying the DNA-protein interactions that are involved in this very first step of horizontal transfer (see next section).

4.3. Protein - Nucleic Acids Interactions

Participants: Isaure Chauvot de Beauchêne [contact person], David Ritchie, Dominique Mias-Lucquin, Antoine Moniot, Honey Jain, Anna Kravchenko, Hrishikesh Dhondge, Malika Smaïl-Tabbone, Marie-Dominique Devignes.

The second application domain of Axis 2 concerns protein-nucleic acids interactions. We need to assess and optimize our new algorithms on concrete protein-nucleic acids complexes in close collaboration with external partners coming from the experimental field of structural biology. To facilitate such collaborations, we will have to create automated and re-usable protein-nucleic acid docking pipelines.

This is the case for our PEPS collaboration "InterANRIL" with the IMoPA lab (CNRS-Université de Lorraine). We are currently working with biologists to apply our fragment-based docking approach to model complexes of the long non-coding RNA (lncRNA) ANRIL with proteins and DNA. In order to extend this approach to partially structured RNA molecules, we have built an automated pipeline to create (i) libraries of RNA fragments with arbitrary characteristics such as secondary structure, and (ii) testing benchmarks for applying these libraries to docking assays.

In the framework of our LUE-FEDER CITRAM project (see above), we adapted this approach and this pipeline to single-strand DNA docking in order to model the complex formed by a bacterial relaxase and its target DNA.

In the future, we will tackle a defined group of RNA-binding proteins containing RNA-Recognition Motif (RRM) domains. We will study existing and predicted complexes between various types of RRMs and various RNA sequences with computational methods in order to calculate CG force-field energy and to help design new synthetic proteins with targeted RNA specificity. This is the goal of the ITN RNAct project and it will require the construction of a dedicated database equipped with querying and analysis facilities, including machine-learning approaches, as well as many interactions within the ITN RNAct consortium.

DYLISS Project-Team

4. Application Domains

4.1. Application fields in biology

In terms of transfer and societal impact, we consider that our role is to develop fruitful collaborations with laboratories of biology in order to consolidate their studies by a smart use of our tools and prototypes and to generate new biological hypotheses to be tested experimentally.

Marine Biology: seaweed enzymes and metabolism & sea-urchin cell-cycle. Our main field of study is **marine biology**, as it is a transversal field covering challenges in integrative biology, dynamical systems and sequence analysis. Our methods based on combinatorial optimization for the reconstruction of genome-scale metabolic networks and on classification of enzyme families based on local and partial alignments allowed the seaweed metabolism *E. Siliculosus* to be deciphered [62], [53]. The study of the *HAD* superfamily of proteins thanks to partial local alignments, produced by *Protomata* tools, allows sub-families to be deciphered and classified, and the metabolic map reconstructed with *Meneco* enabled the reannotation of 56 genes within the *E. siliculosus* genome. These approaches also shed light on evolution of metabolic processes. As a further study, we reconstructed the metabolic network of a symbiot bacterium *Ca. P. ectocarpi* [55] and used this reconstructed network to decipher interactions within the algal-bacteria holobiont, revealing several candidates metabolic pathways for algal-bacterial interactions. Similarily, our analyses suggest that the bacterium *Ca. P. ectocarpi* is able to provide both β-alanine and vitamin B5 to the seaweed via the phosphopantothenate biosynthesis pathway [63].

Micro-biology: elucidating the functioning of extremophile consortiums of bacteria. In this application field, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (co-funded by the Center of Mathematical Modeling, the Center of Regulation Genomics and Inria-Chile). In order to elucidate the main characteristics of these bacteria, our integrative methods were developed to identify the main groups of regulators for their specific response in their living environment. The integrative biology tools *Meneco*, *Lombarde* and *Shogen* have been designed in this context. In particular, genome-scale metabolic network been recently reconstructed and studied with the *Meneco* and *Shogen* approaches, especially on bacteria involved in biomining processes [48] and in Salmon pathogenicity [52].

Agriculture and environmental sciences: upstream controllers of cow, pork and pea-aphid metabolism and regulation. In this application field, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. Our work on the identification of upstream regulators within large-scale knowledge databases (prototype *KeyRegulatorFinder*) [47] and on semantic-based analysis of metabolic networks [45] was very valuable for interpreting differences of gene expression in pork meat [60] and figure out the main gene-regulators of the response of porks to several diets [59]. In addition, constraints-based programming also allows us to decipher regulators of reproduction for the pea aphid, an insect that is a pest on plants. In terms of biological output of the network studies on the pea aphid microRNAs, we have identified one new microRNA (apmir-3019, not present in any known species other than the pea aphid) who has more than 900 putative mRNA targets.

Health: deciphering pathways involved in the TGF-β signalling network. TGF-β is a multifunctional cytokine that regulates mammalian development, differentiation, and homeostasis with both benefical anti-tumor effect [49] and pro-tumor effect [61]. Deciphering protumor versus antitumor signaling requires to take into account a system-wide view and develop predictive models for therapeutic benefit. For that purpose we developed *Cadbiom* and identified gene networks associated with innate immune response to viral infection that combine TGF-β and interleukine signaling pathways [43], [51].

ERABLE Project-Team

4. Application Domains

4.1. Biology and Health

The main areas of application of ERABLE are: (1) biology understood in its more general sense, with a special focus on symbiosis and on intracellular interactions, and (2) health with a special emphasis for now on infectious diseases, rare diseases, and cancer.

GENSCALE Project-Team

4. Application Domains

4.1. Introduction

Today, sequencing data are intensively used in many life science projects. The methodologies developed by the GenScale group are generic approaches that can be applied to a large panel of domains such as health, agronomy or environment areas. The next sections briefly describe examples of our activity in these different domains.

4.2. Health

Genetic and cancer disease diagnostic: Genetic diseases are caused by some particular mutations in the genomes that alter important cell processes. Similarly, cancer comes from changes in the DNA molecules that alter cell behavior, causing uncontrollable growth and malignancy. Pointing out genes with mutations helps in identifying the disease and in prescribing the right drug. Thus, DNA from individual patients is sequenced and the aim is to detect potential mutations that may be linked to the patient disease. Bioinformatics analysis can be based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. One can also scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with genetic or cancer diseases.

Neurodegenerative disorders: The biological processes that lead from abnormal protein accumulation to neuronal loss and cognitive dysfunction is not fully understood. In this context, neuroimaging biomarkers and statistical methods to study large datasets play a pivotal role to better understand the pathophysiology of neurodegenerative disorders. The discovery of new genetic biomarkers could thus have a major impact on clinical trials by allowing inclusion of patients at a very early stage, at which treatments are the most likely to be effective. Correlations with genetic variables can determine subgroups of patients with common anatomical and genetic characteristics.

4.3. Agronomy

Insect genomics: Insects represent major crop pests, justifying the need for control strategies to limit population outbreaks and the dissemination of plant viruses they frequently transmit. Several issues are investigated through the analysis and comparison of their genomes: understanding their phenotypic plasticity such as their reproduction mode changes, identifying the genomic sources of adaptation to their host plant and of ecological speciation, and understanding the relationships with their bacterial symbiotic communities [6].

Improving plant breeding: Such projects aim at identifying favorable alleles at loci contributing to phenotypic variation, characterizing polymorphism at the functional level and providing robust multi-locus SNP-based predictors of the breeding value of agronomical traits under polygenic control. Underlying bioinformatics processing is the detection of informative zones (QTL) on the plant genomes.

4.4. Environment

Food quality control: One way to check food contaminated with bacteria is to extract DNA from a product and identify the different strains it contains. This can now be done quickly with low-cost sequencing technologies such as the MinION sequencer from Oxford Nanopore Technologies.

Ocean biodiversity: The metagenomic analysis of seawater samples provides an original way to study the ecosystems of the oceans. Through the biodiversity analysis of different ocean spots, many biological questions can be addressed, such as the plankton biodiversity and its role, for example, in the CO2 sequestration.

IBIS Project-Team (section vide)

LIFEWARE Project-Team

4. Application Domains

4.1. Preamble

Our collaborative work on biological applications is expected to serve as a basis for groundbreaking advances in cell functioning understanding, cell monitoring and control, and novel therapy design and optimization. Our collaborations with biologists are focused on **concrete biological questions**, and on the building of predictive models of biological systems to answer them. Furthermore, one important application of our research is the development of a **modeling software** for computational systems biology.

4.2. Modeling software for systems biology and synthetic biology

Since 2002, we develop an open-source software environment for modeling and analyzing biochemical reaction systems. This software, called the Biochemical Abstract Machine (BIOCHAM), is compatible with SBML for importing and exporting models from repositories such as BioModels. It can perform a variety of static analyses, specify behaviors in Boolean or quantitative temporal logics, search parameter values satisfying temporal constraints, and make various simulations. While the primary reason of this development effort is to be able to **implement our ideas and experiment them quickly on a large scale**, BIOCHAM is used by other groups either for building models, for comparing techniques, or for teaching (see statistics in software section). BIOCHAM-WEB is a web application which makes it possible to use BIOCHAM without any installation. We plan to continue developing BIOCHAM for these different purposes and improve the software quality.

4.3. Coupled models of the cell cycle and the circadian clock

Recent advances in cancer chronotherapy techniques support the evidence that there exist important links between the cell cycle and the circadian clock genes. One purpose for modeling these links is to better understand how to efficiently target malignant cells depending on the phase of the day and patient characterictics. These questions are at the heart of our collaboration with Franck Delaunay (CNRS Nice) and Francis Lévi (Univ. Warwick, GB, formerly INSERM Hopital Paul Brousse, Villejuif) and of our participation in the ANR HYCLOCK project and in the submitted EU H2020 C2SyM proposal, following the former EU EraNet Sysbio C5SYS and FP6 TEMPO projects. In the past, we developed a coupled model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints [0]. We now focus on the bidirectional coupling between the cell cycle and the circadian clock and expect to gain fundamental insights on this complex coupling from computational modeling and single-cell experiments.

4.4. Biosensor design and implementation in non-living protocells

In collaboration with Franck Molina (CNRS, Sys2Diag, Montpellier) and Jie-Hong Jiang (NTU, Taiwan) we ambition to apply our techniques to the design and implementation of high-level functions in non-living vesicles for medical applications, such as biosensors for medical diagnosis [0]. Our approach is based on purely protein computation and on our ability to compile controllers and programs in biochemical reactions. The realization will be prototyped using a microfluidic device at CNRS Sys2Diag which will allow us to precisely control the size of the vesicles and the concentrations of the injected proteins. It is worth noting that the choice of non-living chassis, in contrast to living cells in synthetic biology, is particularly appealing for security considerations and compliance to forthcoming EU regulation.

[0] Elisabetta De Maria, François Fages, Aurélien Rizk, Sylvain Soliman. Design, Optimization, and Predictions of a Coupled Model of the Cell Cycle, Circadian Clock, DNA Repair System, Irinotecan Metabolism and Exposure Control under Temporal Logic Constraints. Theoretical Computer Science, 412(21):2108 2127, 2011.

[0] Alexis Courbet, Patrick Amar, François Fages, Eric Renard, Franck Molina. Computer-aided biochemical programming of synthetic microreactors as diagnostic devices. Molecular Systems Biology, 14(4), 2018.

4.5. Functional characterization of the resistance of bacterial populations to antimicrobial treatments

Antibiotic resistance is becoming a problem of central importance at a global level. Two mechanisms are at the origin of non-susceptibility to antimicrobial treatments. The first one comes from adaptation of bacterial cells to antibacterial treatments, notably through the modification of efflux pumps or the expression of enzymes that degrade the antibiotics. Cells are individually resistant. The second one, typically found in resistances to β-lactams, a broad class of antibiotics, originates from the release in the environment of the antibiotic degrading enzymes by the dead cells. This leads to population effects by which cells become collectively resilient.

The functional characterization of these different effects is important for the best use of antibiotics (antibiotic stewardship). In collaboration with Lingchong You (Duke University) and with Philippe Glaser (Institut Pasteur), we develop experimental platforms, models, and optimal model calibration methods that gives precise estimations of individual resistance and collective resilience of bacterial populations to antibiotic treatments.

MORPHEME Project-Team (section vide)

MOSAIC Project-Team (section vide)

PLEIADE Project-Team

4. Application Domains

4.1. Genome and transcriptome annotation, to model function

Sequencing genomes and transcriptomes provides a picture of how a biological system can function, or does function under a given physiological condition. Simultaneous sequencing of a group of related organisms is now a routine procedure in biological laboratories for studying a behavior of interest, and provides a marvelous opportunity for building a comprehensive knowledge base of the relations between genomes. Key elements in mining these relations are: classifying the genes in related organisms and the reactions in their metabolic networks, recognizing the patterns that describe shared features, and highlighting specific differences.

PLEIADE will develops applications in comparative genomics of related organisms, using new mathematical tools for representing compactly, at different scales of difference, comparisons between related genomes. New methods based on computational geometry refine these comparisons. Compact representations can be stored, exchanged, and combined. They will form the basis of new simultaneous genome annotation methods, linked directly to abductive inference methods for building functional models of the organisms and their communities.

Our ambition in biotechnology is to permit the design of synthetic or genetically selected organisms at an abstract level, and guide the modification or assembly of a new genome. Our effort is focused on two main applications: genetic engineering and synthetic biology of oil-producing organisms (biofuels in CAER, palm oils), and improving and selecting starter microorganisms used in winemaking (collaboration with the ISVV and the BioLaffort company).

4.1.1. Oil Palm lipid synthesis

The largest source of vegetable oil [0] is the fruit mesocarp of the oil palm *Elaeis guineensis*, a remarkable tissue that can accumulate up to 90% oil, the highest level observed in the plant kingdom. The market share of oil palm is expected to increase in order to meet increased demand for vegetable oil, predicted to double by 2030 [18], be it as food or as a source of biofuels in Africa. A significant proportion of palm oil is produced on small estates that do not have access to efficient milling facilities, and run a great risk of spoilage through oil acidification. Improving palm oil quality through genetics and selection will result in economic gains [24] by addressing several targets such as improvement of oil yield, tuning of oil quality through the rate of unsaturated fatty acids or impairment of degradation processes. Furthermore, as genome biodiversity resides mostly in Africa, oil from African oil palms can vary greatly in fatty acid composition according to cultivar genetic differences and to weather conditions, and the precise mechanisms regulating this variability are not yet understood.

A growing body of molecular resources for studying oil palm fruit are making it possible to study and improve the quality and quantity of oil produced by oil palms. In particular, these oils can vary greatly in fatty acid composition, and while the precise mechanisms regulating this variability are not completely understood, establishing a link between oil palm genotype and phenotype appears increasingly feasible. PLEIADE will work with the CNRS/UB UMR 5200 (LBM), a laboratory with an established reputation in studying fatty acid metabolism in *E. guineensis*, to improve understanding of the links between genetic diversity and oil production, and participate in developing applications.

4.1.2. Engineering pico-algae

[0] 32% of the world market share [24]

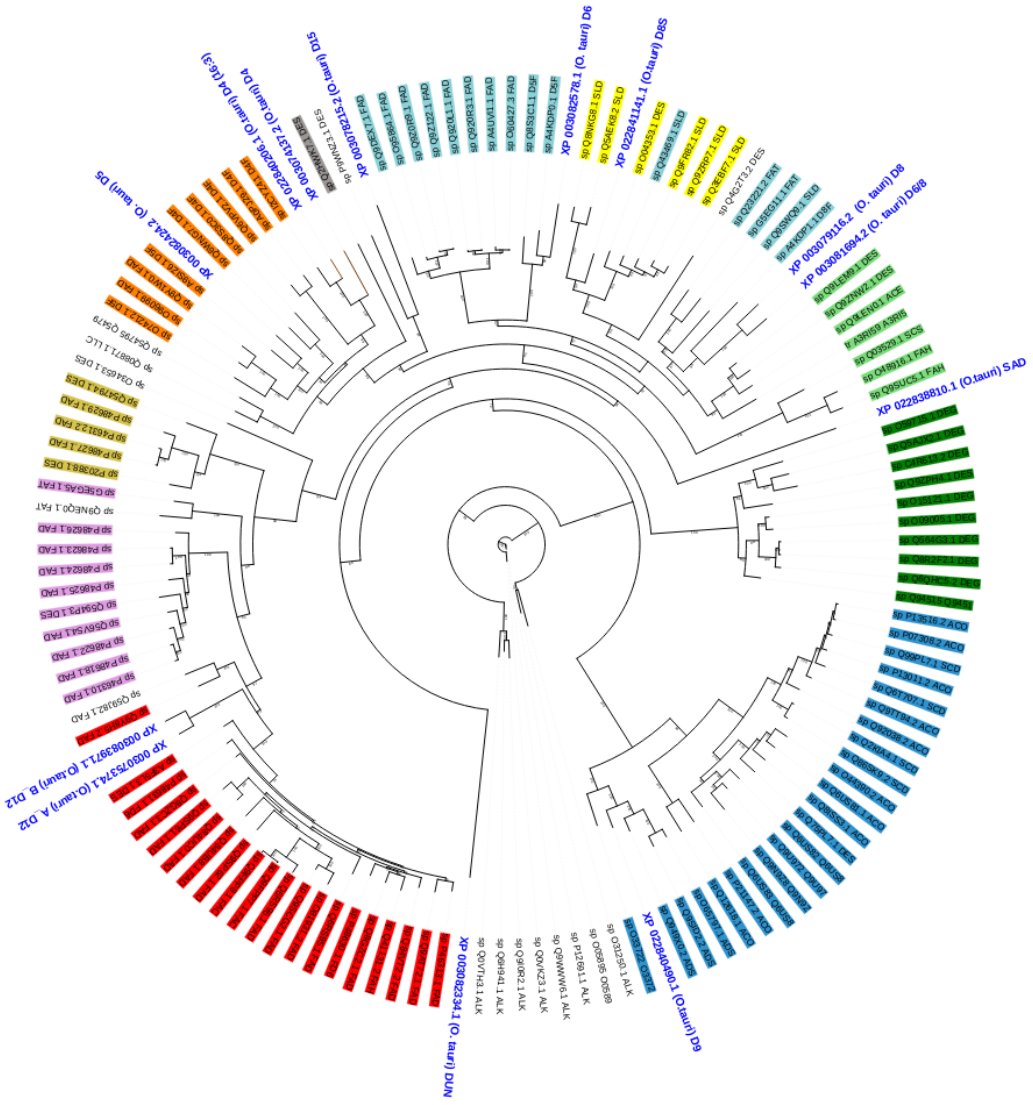

Figure 4. Phylogenetic structure of long-chain polyunsaturated fatty acid desaturase specificity. Highlighted are thirteen desaturases from Ostreococcus tauri

Docosahexaenoic acid (DHA) is an essential nutriment for human brain tissue and can only be obtained from marine or riverine fish that live on phytoplankton and zooplankton, since human neurons lack the delta desaturase required for *de novo* synthesis of DHA. [21] Unfortunately, fishing is become less and less a sustainable resource. Since phytoplankton and zooplankton are the ultimate source of DHA consumed, there is considerable interest in obtaining DHA directly rather than through the intermediary of fish. A very promising approach is through the bio-engineering of pico-algae.

In order to produce the long-chain polyunsaturated fatty acids (LC-PUFA) needed for human nutrition, it is necessary to precisely engineer the desaturases that produce them. Desaturases are enzymes responsible for the introduction of double bonds into fatty acids. Desaturases are specific in recognizing their substrates and in placing the double bond in the proper place. The desaturases that produce the LC-PUFA necessary for human nutrition are present only in some species.

Our goal is to design methods to predict the substrate and region specificities for desaturases in algal species, particularly *Ostreococcus tauri*, the smallest photosynthetic eukaryote that can be cultivated. Thirteen desaturases are known in *O. tauri* and can be placed in the phylogeny of the desaturase family (figure 4). The biochemical and structural characterization of these enzymes is as yet very incomplete. This work is ongoing (see section 8.2.2) and requires close collaboration between biologists and computer scientists.

4.2. Molecular based systematics and taxonomy

Defining and recognizing myriads of species in biosphere has taken phenomenal energy over the past centuries and remains a major goal of Natural History. It is an iconic paradigm in pattern recognition (clustering has coevolved with numerical taxonomy many decades ago). Developments in evolution and molecular biology, as well as in data analysis, have over the past decades enabled a profound revolution, where species can be delimited and recognized by data analysis of sequences. We aim at proposing new tools, in the framework of E-science, which make possible (i) better exploration of the diversity in a given clade, and (ii) assignment of a place in these patterns for new, unknown organisms, using information provided by sets of sequences. This will require investment in data analysis, machine learning, and pattern recognition to deal with the volumes of data and their complexity.

One example of this project is about the diversity of trees in Amazonian forest, in collaboration with botanists in French Guiana. Protists (unicellular Eukaryots) are by far more diverse than plants, and far less known. Molecular exploration of Eukaryotes diversity is nowadays a standard in biodiversity studies. Data are available, through metagenomics, as an avalanche and make molecular diversity enter the domain of Big Data. Hence, an effort will be invested, in collaboration with other Inria teams (GenScale, HiePACS) for porting to HPC algorithms of pattern recognition and machine learning, or distance geometry, for these tools to be available as well in metagenomics. This will be developed first on diatoms (unicellular algae) in collaboration with INRA team at Thonon and University of Uppsala), on pathogens of tomato and grapewine, within an existing network, and on bacterial communities, in collaboration with University of Pau. For the latter, the studies will extend to correlations between molecular diversity and sets of traits and functions in the ecosystem.

4.3. Community ecology and population genetics

Community assembly models how species can assemble or diassemble to build stable or metastable communities. It has grown out of inventories of countable organisms. Using *metagenomics* one can produce molecular based inventories at rates never reached before. Most communities can be understood as pathways of carbon exchange, mostly in the form of sugar, between species. Even a plant cannot exist without carbon exchange with its rhizosphere. Two main routes for carbon exchange have been recognized: predation and parasitism. In predation, interactions–even if sometimes dramatic–may be loose and infrequent, whereas parasitism requires what Claude Combes has called intimate and sustainable interactions [19]. About one decade ago, some works [25] have proposed a comprehensive framework to link the studies of biodiversity with community assembly. This is still incipient research, connecting community ecology and biogeography.

We aim at developping graph-based models of co-occurence between species from NGS inventories in metagenomics, i.e. recognition of patterns in community assembly, and as a further layer to study links, if any, between diversity at different scales and community assemblies, starting from current, but oversimplified theories, where species assemble from a regional pool either randomly, as in neutral models, or by environmental filtering, as in niche modeling. We propose to study community assembly as a multiscale process between nested pools, both in tree communities in Amazonia, and diatom communities in freshwaters. This will be a step towards community genomics, which adds an ecological flavour to metagenomics.

Convergence between the processes that shape genetic diversity and community diversity–drift, selection, mutation/speciation and migration–has been noted for decades and is now a paradigm, establishing a continuous scale between levels of diversity patterns, beyond classical approaches based on iconic levels like species and populations. We will aim at deciphering diversity pattern along these gradients, connecting population and community genetics. Therefore, some key points must be adressed on reliability of tools.

Next-generation sequencing technologies are now an essential tool in population and community genomics, either for making evolutionary inferences or for developing SNPs for population genotyping analyses. Two problems are highlighted in the literature related to the use of those technologies for population genomics: variable sequence coverage and higher sequencing error in comparison to the Sanger sequencing technology. Methods are developed to develop unbiased estimates of key parameters, especially integrating sequencing errors [23]. An additional problem can be created when sequences are mapped on a reference sequence, either the sequenced species or an heterologous one, since paralogous genes are then considered to be the same physical position, creating a false signal of diversity [20]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [20] or by filtering detected SNPs. Finally, an increasingly popular method (RADseq) is used to develop SNP markers, but it was shown that using RADseq data to estimate diversity directly biases estimates [11]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.

SERPICO Project-Team

4. Application Domains

4.1. Modeling and analysis of membrane transport and molecule trafficking at the single cell scale

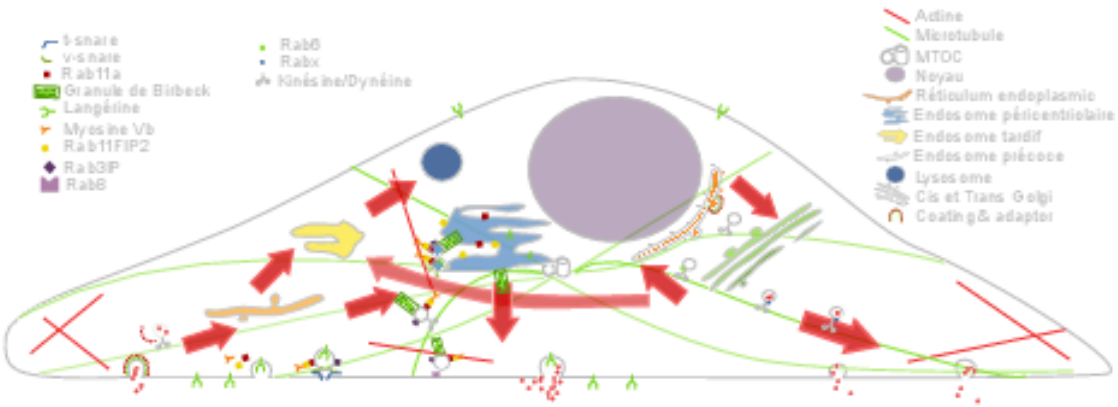

Figure 1. Cargo Langerin Trafficking controlled by Rab11A/Rab11FIP2/MyoVb platform.

In the past recent years, research carried out out together with the "Space Time imaging of Endomembranes and organelles Dynamics" team at CNRS-UMR 144Institut Curie contributed to a better understanding of the intracellular compartmentation, particularly in specialized model cells such as melanocytes and Langerhans cells of the epidermis, of the components and structural events involved in the biogenesis of their specialized organelles: melanosomes and Birbeck granules, respectively and to the understanding on how the dynamics of those structures relate to their physiological functions. These studies have started to highlight: i/ the measurement of multiple sorting and structural events involved in the biogenesis of these organelles; ii/ complexity of the endo-melanosomal network of these highly specialized cells; iii/ complex molecular architecture organizing and coordinating their dynamics; iv/ intracellular transport steps affected in genetic diseases, among which the Hermansky Pudlak syndrome (HPS) or involved in viral infection (HIV and Langerin in Langerhans cells).

In this context, the central aim of SERPICO is to understand how the different machineries of molecular components involved are interconnected and coordinated to generate such specialized structures, an issue that become more and more accessible, thanks to improvement in all domains related to live imaging. We need to address the following topics:

1. developing new bioimaging approaches to observe and statistically analyze such coordinated dynamics in live material;

2. correlating this statistically relevant spatiotemporal organization of protein networks with the biological architectures and at the ultrastructural level;
3. modeling intracellular transport of those reference biological complex systems and proposing new experimental plans in an iterative and virtuous circle;
4. managing and analyzing the workflow of image data obtained along different multidimensional microscopy modalities.

These studies are essential to unravel the complexity of the endomembrane system, how different machineries evolve together and thus coordinate (e.g. see Fig. 1). They help to decipher cell organization and function at different scales through an integrative workflow of methodological and technological developments. New approaches, such as optogenetics may even help controlling cell functions.

At long term, these studies will shed light on the cellular and molecular mechanisms underlying antigen presentation, viral infection or defense mechanisms, skin pigmentation, the pathogenesis of hereditary genetic disorders (lysosomal diseases, immune disorders) and on the mechanisms underlying cell differentiation and cell transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, at a time regime compatible with live cell imaging and close to biochemical molecular interactions. The overview of ultrastructural organization will be achieved by complementary electron microscopy methods which have also undergone a revolutionary improvement over the last decade. Image visualization and quantitative analysis are of course essential issues in this context.

4.2. Imaging and analysis of cytoskeleton dynamics during cell migration

The ability to migrate in space is among the most fundamental functions of eukaryotic cells and thus is one of the best-studied phenomena in biology. During embryonic development, cell movements result in a massive reorganization of the embryo, from a simple spherical ball of cells into a multi-layered organism; many of the cells at or near the surface of the embryo move to a new, more interior location. Moreover, inadequate or inappropriate migration of immune cells is also critically important for the delivery of protective immune responses to tissues and for wound healing. Finally, cell migration may facilitate the dissemination of tumor cells from primary tumor in blood (extravasation) and eventually the colonization of other organs and the formation of secondary tumors.

It has been established that the cytoskeleton, composed of actin filaments, microtubules and intermediate filaments (elongated structures with a diameter of a few dozens of nanometers), is essential for several cell mechanisms, including cell migration, cell division and molecule trafficking:

- i/ the actin filaments promote cell protrusion, adhesion and retraction;
- ii/ the microtubules are the support of molecule traffic and cell polarization;
- iii/ the intermediate filaments are hypothesized to control microtubule organization.

Nevertheless, the mechanical and chemical states of migrating cells under various external conditions remain largely unknown. In the last decade, high-resolution microscopy methods led to the discovery of novel aspects of cell migration. Most approaches and models are limited to migration in 2D, justified by the flatness of the cell-motile mechanisms. However, the mechanical patterns that govern migration in 2D models are often not essential for efficient migration in 3D. Accordingly, recent very challenging 3D models of cells moving on flat surfaces have begun to emerge. The key challenge, however, is to understand how a 3D motile cell crawls through the 3D extracellular matrix. Another issue is of course to measure and understand how membrane protrusion and retraction keep the cell in homeostasis, which of course relate to membrane traffic.

The objective of SERPICO is to develop high-end signal processing and computer vision tools to unfold the dynamical coordination of microtubules, actin filaments and intermediate filaments in 3D, involved in cell migration, cell division and how molecular trafficking is coordinated with cytoskeleton changes in these fundamental cellular functions.

ARAMIS Project-Team

4. Application Domains

4.1. Introduction

We develop different applications of our new methodologies to brain pathologies, mainly neurodegenerative diseases. These applications aim at:

- better understanding the pathophysiology of brain disorders;
- designing systems to support clinical decisions such as diagnosis, prognosis and design of clinical trials;
- developing brain computer interfaces for clinical applications.

4.2. Understanding brain disorders

Computational and statistical approaches have the potential to help understand the pathophysiology of brain disorders. We first aim to contribute to better understand the relationships between pathological processes, anatomical and functional alterations, and symptoms. Moreover, within a single disease, there is an important variability between patients. The models that we develop have the potential to identify more homogeneous disease subtypes, that would constitute more adequate targets for new treatments. Finally, we aim to establish the chronology of the different types of alterations. We focus these activities on neurodegeneratives diseases: dementia (Alzheimer's disease, fronto-temporal dementia), Parkinson's disease, multiple sclerosis.

4.3. Supporting clinical decisions

We aim to design computational tools to support clinical decisions, including diagnosis, prognosis and the design of clinical trials. The differential diagnosis of neurodegenerative diseases can be difficult. Our tools have the potential to help clinicians by providing automated classification that can integrate multiple types of data (clinical/cognitive tests, imaging, biomarkers). Predicting the evolution of disease in individual patients is even more difficult. We aim to develop approaches that can predict which alterations and symptoms will occur and when. Finally, new approaches are needed to select participants in clinical trials. Indeed, it is widely recognized that, to have a chance to be successful, treatments should be administered at a very early stage.

4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the basis of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor patients' progress and their compliance with the protocol, monitoring that they are actually imagining movements. It also follows that feedback from such a BCI can provide patients with an early reinforcement in the critical phase when there is not yet an overt sign of movement recovery.

ATHENA Project-Team

4. Application Domains

4.1. Applications of diffusion MRI

Clinical domain: Diagnosis of neurological disorder
Various examples of CNS diseases as Alzheimer's and Parkinson's diseases and others like multiple sclerosis, traumatic brain injury and schizophrenia have characteristic abnormalities in the micro-structure of brain tissues that are not apparent and cannot be revealed reliably by standard imaging techniques. Diffusion MRI can make visible these co-lateral damages to the fibers of the CNS white matter that connect different brain regions. This is why in our research, Diffusion MRI is the structural imaging modality that will be considered to recover the CNS connectivity.

4.2. Applications of M/EEG

Clinical domain: Diagnosis of neurological disorders
The dream of all M/EEG researchers is to alleviate the need for invasive recordings (electrocorticograms or intracerebral electrodes), which are often necessary prior to brain surgery, in order to precisely locate both pathological and vital functional areas. We are involved in this quest, particularly through our collaborations with the La Timone hospital in Marseille.
Subtopics include:

- Diagnosis of neurological disorders such as epilepsy, schizophrenia, tinnitus, ...
- Presurgical planning of brain surgery.
- Collaboration with the *Institut de Neurosciences des Systèmes* on these topics http://ins.univ-amu.fr/fr/.

Cognitive research

- Aims at better understanding the brain spatio-temporal organisation.
- Collaboration with the *Laboratory for Neurobiology of Cognition* in order to develop methods that suit their needs for sophisticated data analysis.

Brain Computer Interfaces (BCI) aim to allow direct control of external devices using brain signals such as measured through EEG. In our project, BCI can be seen as an application of EEG processing techniques, but also as an object of fundamental and applied research as they open the way for more dynamical and active brain cognitive protocols.

We develop a research collaboration with the eemagine/ANT-Neuro company. We collaborate with Nice University Hospital on the usage of BCI-based communication for ALS[0] patients.

[0]Nice University Hospital hosts a regional reference center for patients suffering from Amyotrophic Lateral Sclerosis

BIOVISION Project-Team

4. Application Domains

4.1. Applications of virtual/augmented reality for low-vision

- **Rehabilitation**: Serious games are games designed for a primary purpose which is not pure entertainment. In our context, we think about serious games as a way to help low-vision patients in performing rehabilitation exercises. Virtual/augmented reality technology is a promising platform to develop such rehabilitation exercises targeted to specific pathologies. For example, with Age Macular Degeneration (AMD), our objective is to propose solutions allowing rehabilitation of visuo-perceptual-motor functions to optimally use residual portions of the peripheral retina and obtain efficient eccentric viewing.
- **Vision aid-systems**: A variety of aids for low-vision people are already on the market using different kinds of virtual/augmented reality platforms (dedicated or large public ones). They offer different functionalities (magnification, image enhancement, text to speech, face and object recognition). Our goal is to design new solutions allowing autonomous interaction in mixed reality environments, and take advantage of the improvement of functions obtained via rehabilitation protocols.
- **Cognitive research**: Virtual/augmented reality technology represents a new opportunity to conduct cognitive and behavioural research using virtual environments where all parameters can be psychophysically controlled. Our objective is to re-assess common theories by allowing patients to freely explore their environment in more ecological conditions.

4.2. Applications of vision modeling studies

- **Neuroscience research.** Making in-silico experiments is a way to reduce the experimental costs, to test hypotheses and design models, and to test algorithms. Our goal is to develop a large-scale simulations platform of impaired retinas, called Macular, allowing to mimic specific degeneracies or pharmacologically induced impairments, as well as to emulate electric stimulation by prostheses. In addition, the platform provides a realistic entry to models or simulators of the thalamus or the visual cortex, in contrast to the entries usually considered in modelling studies.
- **Education.** Macular is also targeted as a useful tool for educational purposes, illustrating for students how the retina works and responds to visual stimuli.

CAMIN Project-Team (section vide)

EMPENN Project-Team

4. Application Domains

4.1. Population imaging

One major objective of neuroimaging researchers and clinicians is to be able to stratify brain imaging data in order to derive new and more specific population models. In practice this requires to set up large-scale experiments that, due to the lack of resources and capabilities to recruit locally subjects who meet specific inclusion criteria, motivates the need for sharing the load.

But, building and using multi-site large-scale resources poses specific challenges to deal with the huge quantity of data produced and their diversity. Empenn will focus on two challenges in particular:

- Provide computational environments for the computation and use of imaging biomarkers in the targeted brain diseases, a solution to be used by radiologists and neurologists/psychiatrists for the clinical follow-up of a large patient population.
- Modeling analytic variability of image processing pipelines to better understand and predict the behaviour of imaging biomarker detection solutions and improve reproducibility and productivity in clinical neuroimaging research.

4.2. Detection and learning

We intend to make significant contributions with major impacts in learning coupling models between functional recordings during neurofeedback procedures. These advances will provide a breakthrough in brain-computer interfaces for rehabilitation protocols. Our aim is to:

- Provide a computational environment that combines data-driven (machine learning) and Bayesian solutions to improve the detection of abnormal patterns in images through decision or evidence theory data fusion strategies. The major initial application will be for multiple sclerosis. Over the longer term, we also expect to adapt these methods to address a wider range of neurological diseases (epilepsy, stroke, tumors, etc.) in neonate and adult brains.
- Develop solutions for combining brain state measurements from multimodal sensors or sequences (e.g. fMRI, ASL, EEG, NIRS, etc.) with applications in the spatiotemporal reconstruction of brain activity from MRI-EEG or the combined detection of the endogenous hemodynamic and resting state network of the brain from ASL and NIRS. Over the longer term, the advent of new hybrid brain imaging sensors (e.g. PET-MRI) will require these methods to be extended to a larger spectrum of information combining structural, morphological, metabolic, electrophysiological and cellular/molecular information (e.g. through the use of specific ligands/nanocarriers).

4.3. Quantitative imaging

The Empenn research group focuses on the development of several quantitative techniques in magnetic resonance imaging of the brain. These methods allow for a characterization of both the function and the structure of the brain with high precision. Arterial spin labelling (ASL) is a contrast agent-free imaging technique which labels arterial blood water as an endogenous tracer for perfusion and can measure resting-state cerebral blood flow. We are interested in estimating multiparametric hemodynamics using ASL, such as combined cerebral blood flow and arterial transit times, and derive statistical descriptors to represent significant differences between groups. In addition to quantitative perfusion parameters, our contributions on tissue compartment imaging aim at delineating neural circuits and characterize their microstructure properties, using both diffusion MRI and relaxometry. In diffusion MRI, arbitrary gradient waveforms were shown to exhibit higher sensitivity to microstructure parameters than standard pulsed gradients. We work on the optimization of sampling protocols in this domain, with the objective to propose sequences compatible with in vivo acquisition.

Complementary to diffusion MRI, we develop methods for the reconstruction of myelin-bound, extra-axonal and cerebrospinal fluid water using multi-compartment modelling of the T2-relaxometry signal. We combine these techniques with tractography to identify trajectories of pathologies associated to the evolution of these microstructural parameters along specific fiber bundles in the brain white matter.

4.4. Behavior

Advances in the field of in vivo imaging offer new opportunities for addressing the management of resistant affective disorders and their consequences (suicide risk and socio-professional impact), and the management of spatial cognition disorders after stroke and their consequences (postural perturbations and the loss of autonomy). Our objective, and the main challenge in this context, will be to introduce medical image computing methods to the multidisciplinary field of behavioral disorders (cognitive disorders, particularly spatial and postural control disorders or anterograde memory impairment, mood disorders, notably resistant depression, schizophrenic disorders, pervasive developmental disorders, attention disorders, etc.) in order to gain a better understanding of the pathology and devise innovative therapeutic approaches.

We also expect to become a major player in the future and make important contributions with significant impacts, primarily in drug-resistant depression in young and old populations. In particular, we expect to provide new image-related metrics combining perfusion, metabolism and microstructural information regarding the brain in order to better characterize pathologies, provide prospective evolution values and potentially provide new brain stimulation targets that could be used in neurofeedback rehabilitation protocols or other types of brain stimulation procedure.

We aim to provide new imaging markers of mental diseases, especially in the context of mood disorders. The new biomarkers will be derived from the metabolic (ASL and later ASL+PET) point of view as well as from the microstructural point of view (multicompartment diffusion MRI and relaxometry). Similarly, we expect to exhibit imaging biomarker regularities combining metabolic and structural information. Over the longer term, we expect these biomarkers to be the target of neurofeedback rehabilitation procedures. Also, over the longer term, we expect to supplement the MRI markers with molecular marker ones coming from new PET tracers, especially those associated with serotonin intake, at one time point or during a rehabilitation protocol under hybrid PET-EEG-MRI neurofeedback procedures.

4.5. Neuroinflammation

Some of the major ongoing research issues in the neuroimaging of neuro-inflammatory diseases concern the definition of new biomarkers for tracking the development of the pathology using high- dimensional data (e.g. nD+t MRI). This includes the use of white matter-specific imaging, such as magnetization transfer MRI, relaxometry and diffusion-weighted imaging (DW-MRI). Our objective is (1) to develop information-processing tools to tag the spatiotemporal evolutions of MS patterns at the brain parenchyma and spinal cord levels from their different signatures (inflammatory cells visible with USPIO or Gd contrast agents on MRI, persistent black holes, eloquent regional atrophy and microstructure signatures); and (2) to test these new tools on new imaging cohorts. In this respect, we for instance conduct studies on brain and spinal cord imaging, continuing on from the PHRC multicentric EMISEP project (PI: G. Edan), as it is very likely that lesions in the spine will directly affect the ambulatory ability of the patient (and thereby the clinical scores). In order to extend this experiment to a larger MS population, based on our expertise from the OFSEP cohort, we also plan to improve the MS therapeutic decision process through the MUSIC project (Multiple Sclerosis Imaging Check out, a public/private project). Our goal is to develop and assess a standardized monitoring tool that provides a robust, long-term computerized MRI follow-up that will become the gold standard in clinical practice for therapeutic decisions in MS treatment. As part of this project, Empenn will share its expertise in data management systems (Shanoir and FLI-IAM) and automatic processing tools (through the medInria and Anima software repositories) to extract quantitative indices from the images.

4.6. Recovery

Mental and neurological disorders are the leading cause of years lived with a disability. Treatment-resistant depression affects approximately 2% of the European population. Meanwhile, in the case of brain disorders, almost 1.5 million Europeans (15 million people worldwide) suffer a stroke event each year. Current recovery methods for brain disorders and traumatic brain injuries remain limited, preventing many from achieving full recuperation. We propose addressing the issue of brain recovery by introducing new advances from recent breakthroughs in computational medical imaging, data processing and human-machine interfaces, and demonstrating how these new concepts can be used, in particular for the treatment of stroke and major depressive disorders.

We ambition to combine advanced instrumental devices (Hybrid EEG, NIRS and MRI platforms), with new hybrid brain computer interface paradigms and new computational models to provide neurofeedback-based therapeutic and neuro-rehabilitation paradigms in some of the major mental and neurological disorders of the developmental and the aging brain.

Neurofeedback involves using a brain-computer interface that provides an individual with real-time biofeedback about his or her brain activity in the form of sensory feedback. It enables individuals to learn to better control their brain activity, which can be measured in real time using various non-invasive sensors as described above. Although EEG is currently the only modality used by clinical practitioners in that context, it lacks specificity due to its low spatial resolution. Dynamic research into fMRI-neurofeedback has held promise for treating depression, chronic pain and stroke, since it offers the prospect of real-time imagery of the activity in deep brain structures with high spatial resolution. However, the low temporal resolution and high cost of fMRI-Neurofeedback has hampered the development of many applications. We believe that the future belongs to hybrid responses that combine multimodal sensors and intend to demonstrate this in the Empenn project.

EPIONE Project-Team (section vide)

MATHNEURO Project-Team (section vide)

MIMESIS Team

4. Application Domains

4.1. Surgical training

Virtual training helps medical students to get familiar with surgical procedures before manipulation of real patients. The development of simulation used for medical training usually requires important computational power, since realistic behaviors are key to deliver a high-fidelity experience to the trainee. Further, the quality of interaction with the simulator (usually via visual and haptic rendering) is also of paramount importance. All these constraints make the development of training systems time-consuming, thus limiting the deployment of virtual simulators in standard medical curriculum.

4.2. Pre-operative planning

Beyond training, clinicians ask for innovative tools that can assist them in the pre-operative planning of an intervention. Using the patient information acquired before the operation, physics-based simulations allow to simulate the effect of therapy with no risk to the patient. The clinicians can thus virtually assess different strategies and select the optimal procedure. Compared to a training simulation, a planning system requires a high accuracy to ensure reliability. Constrained by the time elapsed between the preoperative acquisition and the intervention, the computation must also be efficient.

4.3. Intra-operative guidance

Besides the surgery training and planning, another major need from clinicians is surgical guidance. While the clinician is performing the operation, a guidance system provides enriched visual feedback. This is especially useful with the emergence of minimally invasive surgery (MIS) where the visual information is often strongly limited. It can be used for example to avoid critical areas such as vessels or to highlight the position of a tumor during its resection. In the MIS technique, the clinician does not interact with organs directly as in the open surgery, but manipulates instruments inserted through trocars placed in small incisions in the wall of the abdominal cavity. The surgeon can observe these instruments on a display showing a video stream captured by an endoscopic camera inserted through the navel. The main advantage of the method resides in reducing pain and time recovery, in addition to reducing bleeding and risks of infection. However, from a surgical standpoint, the procedure is quite complex since the field of view is considerably reduced and the direct manipulation of organs is not possible.

MNEMOSYNE Project-Team

4. Application Domains

4.1. Overview

One of the most original specificity of our team is that it is part of a laboratory in Neuroscience (with a large spectrum of activity from the molecule to the behavior), focused on neurodegenerative diseases and consequently working in tight collaboration with the medical domain. As a consequence, neuroscientists and the medical world are considered as the primary end-users of our researches. Beyond data and signal analysis where our expertise in machine learning may be possibly useful, our interactions are mainly centered on the exploitation of our models. They will be classically regarded as a way to validate biological assumptions and to generate new hypotheses to be investigated in the living. Our macroscopic models and their implementation in autonomous robots will allow an analysis at the behavioral level and will propose a systemic framework, the interpretation of which will meet aetiological analysis in the medical domain and interpretation of intelligent behavior in cognitive neuroscience and related domains like for example educational science.

The study of neurodegenerative diseases is targeted because they match the phenomena we model. Particularly, the Parkinson disease results from the death of dopaminergic cells in the basal ganglia, one of the main systems that we are modeling. The Alzheimer disease also results from the loss of neurons, in several cortical and extracortical regions. The variety of these regions, together with large mnesic and cognitive deficits, require a systemic view of the cerebral architecture and associated functions, very consistent with our approach.

Of course, numerical sciences are also impacted by our researches, at several levels. At a global level, we will propose new control architectures aimed at providing a higher degree of autonomy to robots, as well as machine learning algorithms working in more realistic environment. More specifically, our focus on some cognitive functions in closed loop with a real environment will address currently open problems. This is obviously the case for planning and decision making; this is particularly the case for the domain of affective computing, since motivational characteristics arising from the design of an artificial physiology allow to consider not only cold rational cognition but also hot emotional cognition. The association of both kinds of cognition is undoublty an innovative way to create more realistic intelligent systems but also to elaborate more natural interfaces between these systems and human users.

At last, we think that our activities in well-founded distributed computations and high performance computing are not just intended to help us design large scale systems. We also think that we are working here at the core of informatics and, accordingly, that we could transfer some fundamental results in this domain.

NEUROSYS Project-Team

4. Application Domains

4.1. Medical applications

Our research directions are motivated by applications with a high healthcare or social impact. They are developed in collaboration with medical partners, neuroscientists and psychologists. Almost all of our applications can be seen as neural interfaces which require the analysis and modeling of sensorimotor rhythms.

4.1.1. Per-operative awareness during general anesthesia

Collaborators: Univ. Hospital of Nancy-Brabois/dept. Anesthesia & Resuscitation
During general anesthesia, brain oscillations change according to the anesthetic drug concentration. Nowadays, 0.2 to 1.3% of patients regain consciousness during surgery and suffer from post-traumatic disorders. Despite the absence of subject movements due to curare, an electroencephalographic analysis of sensorimotor rhythms can help to detect an intention of movement. Within a clinical protocol, we are working on a brain-computer interface adapted to the detection of intraoperative awareness.

4.1.2. Recovery after stroke

Collaborators: Regional Institute of Physical Medicine and Rehabilitation/Center for Physical Medicine and Rehabilitation (Lay St Christophe), Univ. of Lorraine/PErSEUs.
Stroke is the main cause of acquired disability in adults. Neurosys aims at recovering limb control by improving the kinesthetic motor imagery (KMI) generation of post-stroke patients. We propose to design a KMI-based EEG neural interface which integrates complementary modalities of interactions such as tangible and haptic ones to stimulate the sensorimotor loop. This solution would provide a more engaging and compelling stroke rehabilitation training program based on KMI production.

4.1.3. Modeling Parkinson's disease

Collaborators: Center for Systems Biomedicine (Luxembourg), Institute of Neurodegenerative Diseases (Bordeaux), Human Performance & Robotics laboratory (California State Univ., Long Beach).
Effective treatment of Parkinson's disease should be based on a realistic model of the disease. We are currently developing a neuronal model based on Hodgkin-Huxley neurons reproducing to a certain extent the pathological synchronization observed in basal ganglia in Parkinsonian rats. Moreover, our mesoscopic models of plastic Central Pattern Generator neural circuitries involved in rhythmic movements will allow us to reproduce incoherent coordination of limbs observed on humans affected by Parkinson's diseases like frozen gait, crouch gait. Our long-term objective is to understand how oscillatory activity in the basal ganglia affects motor control in spinal structures.

4.1.4. Modeling propagation of epileptic spikes

Collaborators: Epileptology Unit of the CHRU Nancy (University hospital), CRAN (Research Center in Automation and Signal Processing of Nancy). Effective treatment of patients with refractory epilepsy requires a better understanding of the underlying neuronal mechanisms. In particular, it has been observed that epileptic spikes propagate more easily during stage III sleep (slow wave sleep) than during wakefulness, but the origin of these behaviours still remains misunderstood. At least both, a combination of anatomical structure/connectivity changes and changes in level of neurotransmitters, namely functional connectivity, can cause the propagation. A better knowledge of the functional and structural circuitry could allow a better targetting of structures to be treated, either surgically or pharmacologically, and to better individually adapt the pharmacology to each patient according to their symptomatology.

OPIS Project-Team

4. Application Domains

4.1. Sparse signal processing in chemistry

Participants: Marc Castella, Emilie Chouzenoux, Arthur Marmin, Jean-Christophe Pesquet (Collaboration: Laurent Duval, IFPEN, Rueil Malmaison)
Mass Spectrometry (MS) is a powerful tool used for robust, accurate, and sensitive detection and quantification of molecules of interest. Thanks to its sensibility and selectivity, MS is widely used in proteomics such anti-doping, metabolomics, medicine or structural biology. In particular, it has applications in clinical research, personalized medicine, diagnosis process and tumours profiling and pharmaceutical quality control. In an MS experiment, the raw signal arising from the molecule ionization in an ion beam is measured as a function of time via Fourier Transform-based measures such as Ion Cyclotron Resonance (FT-ICR) and Orbitrap. A spectral analysis step is then performed to improve the quality of data. The goal is then to determine from this observed pattern distribution the most probable chemical composition of the sample, through the determination of the monoisotopic mass, charge state and abundance of each present molecule. This amounts to solve a large scale signal estimation problem under specific sparsity constraints [35], [55]. Collaboration with Dr. L. Duval, Research Engineer at IFP Energies Nouvelles, France is on-going in this applicative context.

4.2. Image restoration for two-photon microscopy

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Mathieu Chalvidal (Collaboration: Claire Lefort, XLIM, CNRS, Limoges)
Through an ongoing collaboration with physicists from XLIM laboratory (CNRS, Limoges, France), we propose advanced mathematical and computational solutions for multiphoton microscopy (MPM) 3D image restoration. This modality enjoys many benefits such as a decrease in phototoxicity and increase in penetration depth. However, blur and noise issues can be more severe than with standard confocal images. Our objective is to drastically improve the quality of the generated images and their resolution by improving the characterization of the PSF of the system [12] and compensating its effect. We consider the application of the improved MPM imaging tool to the microscopic analysis of muscle ultrastructure and composition, with the aim to help diagnosing muscle disorders including rare and orphan muscle pathologies.

4.3. Representation Learning for Biological Networks

Participants: Fragkiskos Malliaros, Abdulkadir Çelikkanat (Collaboration: Duong Nguyen, UC San Diego)
Networks (or graphs) are ubiquitous in the domain of biology, as many biological systems can naturally be mapped to graph structures. Characteristic examples include protein-protein interaction and gene regulatory networks. To this extend, machine learning on graphs is an important task with many practical applications in network biology. For example, in the case on protein-protein interaction networks, predicting the function of a protein is a key task that assigns biochemical roles to proteins. The main challenge here is to find appropriate representations of the graph structure, in order to be easily exploited by machine learning models. The traditional approach to the problem was relying on the extraction of “hand-crafted" discriminating features that encode information about the graph, based on user-defined heuristics. Nevertheless, this approach has demonstrated severe limitations, as the learning process heavily depends on the manually extracted features. To this end, feature (or representation) learning techniques can be used to automatically learn to encode the graph structure into low-dimensional feature vectors – which can later be used in learning tasks. Our goal here is to develop a systematic framework for large-scale representation learning on biological graphs. Our approach takes advantage of the clustering structure of these networks, to further enhance the ability of the learned features to capture intrinsic structural properties.

4.4. Breast tomosynthesis

Participants: Emilie Chouzenoux, Jean-Christophe Pesquet, Maissa Sghaier (collaboration G. Palma, GE Healthcare)

Breast cancer is the most frequently diagnosed cancer for women. Mammography is the most used imagery tool for detecting and diagnosing this type of cancer. Since it consists of a 2D projection method, this technique is sensitive to geometrical limitations such as the superimposition of tissues which may reduce the visibility of lesions or make even appear false structures which are interpreted by radiologists as suspicious signs. Digital breast tomosynthesis allows these limitations to be circumvented. This technique is grounded on the acquisition of a set of projections with a limited angle view. Then, a 3D estimation of the sensed object is performed from this set of projections, so reducing the overlap of structures and improving the visibility and detectability of lesions possibly present in the breast. The objective of our work is to develop a high quality reconstruction methodology where the full pipeline of data processing is modeled [50].

4.5. Inference of gene regulatory networks

Participants: Surabhi Jagtap, Fragkiskos Malliaros, Jean-Christophe Pesquet (collaboration A. Pirayre and L. Duval, IFPEN)

The discovery of novel gene regulatory processes improves the understanding of cell phenotypic responses to external stimuli for many biological applications, such as medicine, environment or biotechnologies. To this purpose, transcriptomic data are generated and analyzed from DNA microarrays or more recently RNAseq experiments. They consist in genetic expression level sequences obtained for all genes of a studied organism placed in dierent living conditions. From these data, gene regulation mechanisms can be recovered by revealing topological links encoded in graphs. In regulatory graphs, nodes correspond to genes. A link between two nodes is identified if a regulation relationship exists between the two corresponding genes. In our work, we propose to address this network inference problem with recently developed techniques pertaining to graph optimization. Given all the pairwise gene regulation information available, we propose to determine the presence of edges in the considered GRN by adopting an energy optimization formulation integrating additional constraints. Either biological (information about gene interactions) or structural (information about node connectivity) a priori are considered to restrict the space of possible solutions. Different priors lead to different properties of the global cost function, for which various optimization strategies, either discrete and continuous, can be applied.

4.6. Imaging biomarkers and characterization for chronic lung diseases

Participants: Guillaume Chassagnon, Maria Vakalopoulou (in collaboration with Marie-Pierre Revel and Nikos Paragios: AP-HP - Hopital Cochin Broca Hotel Dieu; Therapanacea)

Diagnosis and staging of chronic lung diseases is a major challenge for both patient care and approval of new treatments. Among imaging techniques, computed tomography (CT) is the gold standard for in vivo morphological assessment of lung parenchyma currently offering the highest spatial resolution in chronic lung diseases. Although CT is widely used its optimal use in clinical practice and as an endpoint in clinical trials remains controversial. Our goal is to develop quantitative imaging biomarkers that allow (i)severity assessment (based on the correlation to functional and clinical data) and (ii) monitoring the disease progression. In the current analysis we focus on scleroderma and cystic fibrosis as models for restrictive and obstructive lung disease, respectively. Two different approaches are investigated: disease assessment by deep convolutional neural networks and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

4.7. Imaging radiomics and genes to assess immunotherapy

Participants: Samy Ammari, Enzo Batistella, Emilie Chouzenoux, Théo Estienne, Marvin Lerousseau, Hugues Talbot, Roger Sun, Maria Vakalopoulou (in collaboration with Corinne Balleyguier, Caroline Caramella, Éric Deutsch, Nathalie Lassau, Institut de Cancérologie Gustave Roussy, Nikos Paragios, Therapanacea)

Because responses of cancer patients to immunotherapy can vary considerably, innovative predictors of response to treatment are urgently needed to improve patients outcomes.

We have aimed to develop and independently validate a radiomics-based biomarkers of tumour-infiltrating CD8 cells in patients included in phase 1 trials of anti-programmed cell death protein (PD)-1 or anti-programmed cell death ligand 1 (PD-L1) mono-therapy. We also aimed to evaluate the association between the biomarker, tumour immune phenotype and clinical outcomes of these patients.

Concurrently, we have evaluated various ways of estimating patient response to treatment based on well-established radiomics such as estimated tumour count and volumes. Among published metrics, we have select those that shown good predictive power and proposed a new one, which is particularly effective for patient with a poor response [63].

Furthermore, we have developed and validated a novel imaging-based decision-making algorithm for use by the clinician that helps differentiate pituitary metastasis from autoimmune hypophysitis in patients undergoing immune checkpoint blockade therapy [21].

These works are in collaboration with the Institut de Cance´rologie Gustave Roussy Paris.

4.8. Development of a heart ventricle vessel generation model for perfusion analysis

Participant: Hugues Talbot (collaboration with L. Najman ESIEE Paris, I. Vignon-Clementel, REO Team leader, Inria, Charles Taylor, Heartflow Inc.)
Cardio-vascular diseases are the leading cause of mortality in the world. Understanding these diseases is a current, challenging and essential research project. The leading cause of heart malfunction are stenoses causing ischemia in the coronary vessels. Current CT and MRI technology can assess coronary diseases but are typically invasive, requiring catheterization and relatively toxic contrast agents injection. In collaboration with the REO team headed by Irène Vignon-Clementel, and Heartflow, a US based company, we have in the past worked to use image-based exams only, limiting the use of contrast agents and in many cases eliminating catheterisation. Heartflow is current the market leader in non-invasive coronary exams.

Unfortunately, current imaging technology is unable to assess the full length of coronary vessels. CT is limited to a resolution of about 1mm, whereas coronary vessels can be much smaller, down to about 10 micrometers in diameter. Blood perfusion throughout the heart muscle can provide insight regarding coronary health in areas that CT or MRI cannot assess. Perfusion imaging with PET or a Gamma camera, the current gold standard, is an invasive technology requiring the use of radioactive tracers.

We have investigated patient-specific vessel generation models together with porous model simulations in order to propose a forward model of perfusion imaging, based on the known patient data, computer flow dynamic simulations as well as experimental data consistent with known vessel and heart muscle physiology. The objective of this work is to both provide a useful, complex forward model of perfusion image generation, and to solve the inverse problem of locating and assessing coronary diseases given a perfusion exam, even though the affected vessels may be too small to be imaged directly.

In 2019, we have produced a functional myocardial perfusion model consisting of the CT-derived segmented coronary vessels, a simulated vessel tree consisting of several thousands of terminal vessels, filling the myocardium in a patient-specific way, consistent with physiology data, physics-based and empirically-observed vessel growth rules, and a porous medium. We have produced a CFD code capable of simulating blood flow in all three coupled compartments, which allows us to simulate perfusion realistically.

PARIETAL Project-Team

4. Application Domains

4.1. Cognitive neuroscience

4.1.1. Macroscopic Functional cartography with functional Magnetic Resonance Imaging (fMRI)

The brain as a highly structured organ, with both functional specialization and a complex network organization. While most of the knowledge historically comes from lesion studies and animal electophysiological recordings, the development of non-invasive imaging modalities, such as fMRI, has made it possible to study routinely high-level cognition in humans since the early 90's. This has opened major questions on the interplay between mind and brain , such as: How is the function of cortical territories constrained by anatomy (connectivity) ? How to assess the specificity of brain regions ? How can one characterize reliably inter-subject differences ?

4.1.2. Analysis of brain Connectivity

Functional connectivity is defined as the interaction structure that underlies brain function. Since the beginning of fMRI, it has been observed that remote regions sustain high correlation in their spontaneous activity, i.e. in the absence of a driving task. This means that the signals observed during resting-state define a signature of the connectivity of brain regions. The main interest of resting-state fMRI is that it provides easy-to-acquire functional markers that have recently been proved to be very powerful for population studies.

4.1.3. Modeling of brain processes (MEG)

While fMRI has been very useful in defining the function of regions at the mm scale, Magneto-encephalography (MEG) provides the other piece of the puzzle, namely temporal dynamics of brain activity, at the ms scale. MEG is also non-invasive. It makes it possible to keep track of precise schedule of mental operations and their interactions. It also opens the way toward a study of the rhythmic activity of the brain. On the other hand, the localization of brain activity with MEG entails the solution of a hard inverse problem.

4.1.4. Current challenges in human neuroimaging (acquisition+analysis)

Human neuroimaging targets two major goals: *i)* the study of neural responses involved in sensory, motor or cognitive functions, in relation to models from cognitive psychology, i.e. the identification of neurophysiological and neuroanatomical correlates of cognition; *ii)* the identification of markers in brain structure and function of neurological or psychiatric diseases. Both goals have to deal with a tension between

- the search for higher spatial [0] resolution to increase **spatial specificity** of brain signals, and clarify the nature (function and structure) of brain regions. This motivates efforts for high-field imaging and more efficient acquisitions, such as compressed sensing schemes, as well as better source localization methods from M/EEG data.
- the importance of inferring brain features with **population-level** validity, hence, contaminated with high variability within observed cohorts, which blurs the information at the population level and ultimately limits the spatial resolution of these observations.

[0] and to some extent, temporal, but for the sake of simplicity we focus here on spatial aspects.

Importantly, the signal-to-noise ratio (SNR) of the data remains limited due to both resolution improvements [0] and between-subject variability. Altogether, these factors have led to realize that results of neuroimaging studies were **statistically weak**, i.e. plagued with low power and leading to unreliable inference [72], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [73]): this is at the core of the *neuroimaging reproducibility crisis*. This crisis is deeply related to a second issue, namely that only few neuroimaging datasets are publicly available, making it impossible to re-assess a posteriori the information conveyed by the data. Fortunately, the situation improves, lead by projects such as NeuroVault or OpenfMRI. A framework for integrating such datasets is however still missing.

[0]The SNR of the acquired signal is proportional to the voxel size, hence an improvement by a factor of 2 in image resolution along each dimension is payed by a factor of 8 in terms of SNR.

AIRSEA Project-Team

4. Application Domains

4.1. The Ocean-Atmosphere System

The evolution of natural systems, in the short, mid, or long term, has extremely important consequences for both the global Earth system and humanity. Forecasting this evolution is thus a major challenge from the scientific, economic, and human viewpoints.

Humanity has to face the problem of **global warming**, brought on by the emission of greenhouse gases from human activities. This warming will probably cause huge changes at global and regional scales, in terms of climate, vegetation and biodiversity, with major consequences for local populations. Research has therefore been conducted over the past 15 to 20 years in an effort to model the Earth's climate and forecast its evolution in the 21st century in response to anthropic action.

With regard to short-term forecasts, the best and oldest example is of course **weather forecasting**. Meteorological services have been providing daily short-term forecasts for several decades which are of crucial importance for numerous human activities.

Numerous other problems can also be mentioned, like **seasonal weather forecasting** (to enable powerful phenomena like an El Niño event or a drought period to be anticipated a few months in advance), **operational oceanography** (short-term forecasts of the evolution of the ocean system to provide services for the fishing industry, ship routing, defense, or the fight against marine pollution) or the prediction of **floods**.

As mentioned previously, mathematical and numerical tools are omnipresent and play a fundamental role in these areas of research. In this context, the vocation of AIRSEA is not to carry out numerical prediction, but to address mathematical issues raised by the development of prediction systems for these application fields, in close collaboration with geophysicists.

ANGE Project-Team

4. Application Domains

4.1. Overview

Sustainable development and environment preservation have a growing importance and scientists have to address difficult issues such as: management of water resources, renewable energy production, bio/geo-chemistry of oceans, resilience of society w.r.t. hazardous flows, urban pollutions, ...

As mentioned above, the main issue is to propose models of reduced complexity, suitable for scientific computing and endowed with stability properties (continuous and/or discrete). In addition, models and their numerical approximations have to be confronted with experimental data, as analytical solutions are hardly accessible for these problems/models. A. Mangeney (IPGP) and N. Goutal (EDF) may provide useful data.

4.2. Geophysical flows

Reduced models like the shallow water equations are particularly well-adapted to the modelling of geophysical flows since there are characterized by large time or/and space scales. For long time simulations, the preservation of equilibria is essential as global solutions are a perturbation around them. The analysis and the numerical preservation of non-trivial equilibria, more precisely when the velocity does not vanish, are still a challenge. In the fields of oceanography and meteorology, the numerical preservation of the so-called geostrophic state, which is the balance between the gravity field and the Coriolis force, can significantly improve the forecasts. In addition, data assimilation is required to improve the simulations and correct the dissipative effect of the numerical scheme.

The sediment transport modelling is of major interest in terms of applications, in particular to estimate the sustainability of facilities with silt or scour, such as canals and bridges. Dredging or filling-up operations are expensive and generally not efficient in the long term. The objective is to determine a configuration almost stable for the facilities. In addition, it is also important to determine the impact of major events like emptying dam which is aimed at evacuating the sediments in the dam reservoir and requires a large discharge. However, the downstream impact should be measured in terms of turbidity, river morphology and flood.

4.3. Hydrological disasters

It is a violent, sudden and destructive flow. Between 1996 and 2005, nearly 80% of natural disasters in the world have meteorological or hydrological origines. The main interest of their study is to predict the areas in which they may occur most probably and to prevent damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worst damages. For example, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last require the use of reduced models like the shallow water equations. In urban areas, the flow can be largely impacted by the debris, in particular cars, and this requires fluid/structure interactions be well understood. Moreover, underground flows, in particular in sewers, can accelerate and amplify the flow. To take them into account, the model and the numerical resolution should be able to treat the transition between free surface and underground flows.

Tsunamis are another hydrological disaster largely studied. Even if the propagation of the wave is globally well described by the shallow water model in oceans, it is no longer the case close to the epicenter and in the coastal zone where the bathymetry leads to vertical accretions and produces substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

While viscous effects can often be neglected in water flows, they have to be taken into account in situations such as avalanches, debris flows, pyroclastic flows, erosion processes, *...i.e.* when the fluid rheology becomes more complex. Gravity driven granular flows consist of solid particles commonly mixed with an interstitial lighter fluid (liquid or gas) that may interact with the grains and decrease the intensity of their contacts, thus reducing energy dissipation and favoring propagation. Examples include subaerial or subaqueous rock avalanches (e.g. landslides).

4.4. Biodiversity and culture

Nowadays, simulations of the hydrodynamic regime of a river, a lake or an estuary, are not restricted to the determination of the water depth and the fluid velocity. They have to predict the distribution and evolution of external quantities such as pollutants, biological species or sediment concentration.

The potential of micro-algae as a source of biofuel and as a technological solution for CO2 fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds. One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps, ...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel). Consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs. Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed. However, it is clear that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature and density, evaporation and external forcing.

Many problems previously mentioned also arise in larger scale systems like lakes. Hydrodynamics of lakes is mainly governed by geophysical forcing terms: wind, temperature variations, ...

4.5. Sustainable energy

One of the booming lines of business is the field of renewable and decarbonated energies. In particular in the marine realm, several processes have been proposed in order to produce electricity thanks to the recovering of wave, tidal and current energies. We may mention water-turbines, buoys turning variations of the water height into electricity or turbines motioned by currents. Although these processes produce an amount of energy which is less substantial than in thermal or nuclear power plants, they have smaller dimensions and can be set up more easily.

The fluid energy has kinetic and potential parts. The buoys use the potential energy whereas the water-turbines are activated by currents. To become economically relevant, these systems need to be optimized in order to improve their productivity. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy.

This is a complex and original issue which requires a fine model of energy exchanges and efficient numerical tools. In a second step, the optimisation of parameters that can be changed in real-life, such as bottom bathymetry and buoy shape, must be studied. Eventually, physical experiments will be necessary for the validation.

4.6. Urban environment

The urban environment is essentially studied for air and noise pollutions. Air pollution levels and noise pollution levels vary a lot from one street to next. The simulations are therefore carried out at street resolution and take into account the city geometry. The associated numerical models are subject to large uncertainties. Their input parameters, e.g. pollution emissions from road traffic, are also uncertain. Quantifying

the simulation uncertainties is challenging because of the high computational costs of the numerical models. An appealing approach in this context is the use of metamodels, from which ensembles of simulations can be generated for uncertainty quantification.

The simulation uncertainties can be reduced by the assimilation of fixed and mobile sensors. High-quality fixed monitoring sensors are deployed in cities, and an increasing number of mobile sensors are added to the observational networks. Even smartphones can be used as noise sensors and dramatically increase the spatial coverage of the observations. The processing and assimilation of the observations raises many questions regarding the quality of the measurements and the design of the network of sensors.

4.7. SmartCity

There is a growing interest for environmental problems at city scale, where a large part of the population is concentrated and where major pollutions can occur. Numerical simulation is well established to study the urban environment, *e.g.* for road traffic modelling. As part of the smartcity movement, an increasing number of sensors collect measurements, at traditional fixed observation stations, but also on mobile devices, like smartphones. They must properly be taken into account given their number but also their potential low quality.

Pratical applications include air pollution and noise pollution. These directly relate to road traffic. Data assimilation and uncertainty propagation are key topics in these applications.

CASTOR Project-Team (section vide)

COFFEE Project-Team

4. Application Domains

4.1. Multiphase porous media flows and multi-physics coupling

Our research focuses on the numerical modeling of multiphase porous media flows accounting for complex geology and for nonlinear and multi-physics couplings. It is applied to various problems in the field of energy such as the simulation of geothermal systems in collaboration with BRGM, of nuclear waste repositories in collaboration with Andra, and of oil and gas recovery in collaboration with Total. Our research directions include the development of advanced numerical schemes adapted to polyhedral meshes and highly heterogeneous media in order to represent more accurately complex geologies. A special focus is made on the modeling of multiphase flows in network of faults or fractures represented as interfaces of co-dimension one coupled to the surrounding matrix. We also investigate nonlinear solvers adapted to the nonlinear couplings between gravity, capillary and viscous forces in highly heterogeneous porous media. In the same line, we study new domain decomposition algorithms to couple non-isothermal compositional liquid gas flows in a porous medium with free gas flows occurring at the interface between the ventilation gallery and the nuclear waste repository or between a geothermal reservoir and the atmosphere. We have begun exploring the coupling between the multiphase flow in the porous matrix and the solid mechanics involved in opening fractures.

4.2. Particulate and mixture flows

We investigate fluid mechanics models referred to as “multi–fluids” flows. A large part of our activity is more specifically concerned with the case where a disperse phase interacts with a dense phase. Such flows arise in numerous applications, like for pollutant transport and dispersion, the combustion of fuel particles in air, the modelling of fluidized beds, the dynamic of sprays and in particular biosprays with medical applications, engine fine particles emission... There are many possible modelings of such flows: microscopic models where the two phases occupy distinct domains and where the coupling arises through intricate interface conditions; macroscopic models which are of hydrodynamic (multiphase) type, involving non standard state laws, possibly with non conservative terms, and the so–called mesoscopic models. The latter are based on Eulerian–Lagrangian description where the disperse phase is described by a particle distribution function in phase space. Following this path we are led to a Vlasov-like equation coupled to a system describing the evolution of the dense phase that is either the Euler or the Navier-Stokes equations. It turns out that the leading effect in such models is the drag force. However, the role of other terms, of more or less phenomenological nature, deserves to be discussed (close packing terms, lift term, Basset force...). Of course the fluid/kinetic model is interesting in itself and needs further analysis and dedicated numerical schemes. In particular, in collaboration with the Atomic Energy Commission (CEA), we have proposed a semi-Lagrangian scheme for the simulation of particulate flows, extending the framework established in plasma physics to such flows.

We also think it is worthwhile to identify hydrodynamic regimes: it leads to discuss hierarchies of coupled hydrodynamic systems, the nature of which could be quite intriguing and original, while they share some common features of the porous media problems. We are particularly interested in revisiting the modeling of mixture flows through the viewpoint of kinetic models and hydrodynamic regimes. We propose to revisit the derivation of new mixture models, generalizing Kazhikov-Smagulov equations, through hydrodynamic asymptotics. The model is of “hybrid” type in the sense that the constraint reduces to the standard incompressibility condition when the disperse phase is absent, while it involves derivatives of the particle volume fraction when the disperse phase is present.

4.3. Fungal network growth

Members of the team have started an original research program devoted to fungal network growth. We started working on this subject through a collaboration with biologists and physicists at LIED (Université Paris Diderot) and probabilists in CMAP (Ecole Polytechnique) and Université Paris Sud, involving Rémi Catellier and Yves D'Angelo in LJAD in Nice. The motivation is to understand branching networks as an efficient space exploration strategy, with fungus *Podospora Anserina* being the biological model considered. This research is submitted as an ANR-project and has been supported by various local fundings.

FLUMINANCE Project-Team (section vide)

LEMON Project-Team (section vide)

MAGIQUE-3D Project-Team

4. Application Domains

4.1. Seismic Imaging

The main objective of modern seismic processing is to find the best representation of the subsurface that can fit the data recorded during the seismic acquisition survey. In this context, the seismic wave equation is the most appropriate mathematical model. Numerous research programs and related publications have been devoted to this equation. An acoustic representation is suitable if the waves propagate in a fluid. But the subsurface does not contain fluids only and the acoustic representation is not sufficient in the general case. Indeed the acoustic wave equation does not take some waves into account, for instance shear waves, turning waves or the multiples that are generated after several reflections at the interfaces between the different layers of the geological model. It is then necessary to consider a mathematical model that is more complex and resolution techniques that can model such waves. The elastic or viscoelastic wave equations are then reference models, but they are much more difficult to solve, in particular in the 3D case. Hence, we need to develop new high-performance approximation methods.

Reflection seismics is an indirect measurement technique that consists in recording echoes produced by the propagation of a seismic wave in a geological model. This wave is created artificially during seismic acquisition surveys. These echoes (i.e., reflections) are generated by the heterogeneities of the model. For instance, if the seismic wave propagates from a clay layer to sand, one will observe a sharp reflected signal in the seismic data recorded in the field. One then talks about reflection seismics if the wave is reflected at the interface between the two media, or talks about seismic refraction if the wave is transmitted along the interface. The arrival time of the echo enables one to locate the position of this transition, and the amplitude of the echo gives information on some physical parameters of the two geological media that are in contact. The first petroleum exploration surveys were performed at the beginning of the 1920's and for instance, the Orchard Salt Dome in Texas (USA) was discovered in 1924 by the seismic-reflection method.

4.2. Imaging complex media with ultrasonic waves

The acoustic behavior of heterogeneous or composite materials attracts considerable excitement. Indeed, their acoustic response may be extremely different from the single constituents responses. In particular, dispersions of resonators in a matrix are the object of large research efforts, both experimentally and theoretically. However it is still a challenge to dispose of numerical tools with sufficient abilities to deal with the simulation and imaging of such materials behavior. Indeed, not only acoustic simulations are very time-consuming, but they have to be performed on realistic enough solution domains, i.e. domains which capture well enough the structural features of the considered materials.

This collaboration with I2M, University of Bordeaux aims at addressing this type of challenges by developing numerical and experimental tools in order to understand the propagation of ultrasonic waves in complex media, image these media, and in the future, help design composite materials for industrial purposes.

4.3. Helioseismology

This collaboration with the Max Planck Institute for Solar System, Göttingen, Germany, which started in 2014, aims at designing efficient numerical methods for the wave propagation problems that arise in helioseismology in the context of inverse problems. The final goal is to retrieve information about the structure of the Sun i.e. inner properties such as density or pressure via the inversion of a wave propagation problem. Acoustic waves propagate inside the Sun which, in a first approximation and regarding the time scales of physical phenomena, can be considered as a moving fluid medium with constant velocity of motion. Some other simplifications lead to computational saving, such as supposing a radial or axisymmetric geometry of the Sun.

Aeroacoustic equations must be adapted and efficiently solved in this context, this has been done in the finite elements code Montjoie. In other situations, a full 3D simulation is required and demands large computational resources. Ultimately, we aim at modeling the coupling with gravity potential and electromagnetic waves (MHD equations) in order to be able to better understand Sun spots.

SERENA Project-Team

4. Application Domains

4.1. Multiphase flows and transport of contaminants in the subsurface

- subsurface depollution after chemical leakage
- nuclear waste disposal in deep underground repositories
- flow in large scale discrete fracture networks
- production of oil and gas

4.2. Industrial risks in energy production

- Stokes and Navier–Stokes flows related to nuclear reactor operation
- seismic wave propagation for detection and protection
- electromagnetism for interfaces between dielectrics and negative metamaterials

4.3. Nonlinear mechanics

- quasi-static and dynamic elastoplastic evolutions with small and large deformations
- quasi-static and dynamic crack propagation
- nonlinear contact and friction conditions
- application to engineering components mainly related to nuclear reactor operation and safety analysis

4.4. Computational quantum chemistry

- guaranteed bounds for ground-state energy (eigenvalues) and ground-state density matrix (eigenvectors) in first-principle molecular simulation
- application to Laplace, Gross–Pitaevskii, Kohn–Sham, and Schrödinger models

STEEP Project-Team

4. Application Domains

4.1. Introduction

In the context described in the previous sections, we can distinguish two connected and complementary strategies for analyzing environmental pressures: a sectorial approach and a spatial one. The first one is more directly connected to ecological accounting, the second one has more direct relations to urban economy and land cover modelling. We start by describing the former.

4.2. Ecological accounting for sectorial pressure assessment

One of the major issues in the assessment of the long-term sustainability of urban areas is related to the concept of "imported sustainability". Cities bring in from the outside most of their material and energy resources, and reject to the outside the waste produced by their activity. The modern era has seen a dramatic increase in both volume and variety of these material flows and consumption as well as in distance of origin and destination of these flows, usually accompanied by a spectacular increase in the associated environmental impacts. A realistic assessment of the sustainability of urban areas requires to quantify both local and distant environmental impacts; greenhouse gas emissions are only one aspect of this question. Such an assessment brings to light the most relevant direct and indirect lines of action on these issues. In this respect, it is useful to introduce the alternative concepts of consumer versus producer responsibility (or point of view).

The producer point of view is the most useful to pinpoint relevant direct lines of actions on environmental pressures due to production. In other respects, any territory imports and exports goods and services from and to the rest of the world. The consumer point of view provides information on the indirect pressures associated with these exchanges, as production responds to a final demand. Tracking the various supply chains through the analysis of the structure of the local economy and its relations and dependencies to the external world allows us to identify critically important contributions to environmental pressures; this also enables us to define fair environmental indicators in order not to attribute environmental pressures to producers only (whose responsibility is the easier to quantify of the two). In this approach, the producer responsibility follows directly from the measurement of its energy and material uses, while the consumer responsibility is established indirectly through an allocation of the impacts of production to the final consumers, but this second mode of allocation is to some extent virtual and partly subjective. Four methods stand out:

- Material Flow Analysis (MFA)
- Input-Output Analysis (IOA)
- Life-Cycle Analysis (LCA)
- Ecological Footprint (EF)

Each of these is based on a well-defined structuring element: mass conservation for MFA, measure of industrial inter-dependencies for IOA, identification of all the steps from cradle to grave for LCA, measure of biocapacity demand for EF. The different methods have preferred areas of application. For example, EF is more relevant for analyzing primary production such as agricultural staples, wood, etc. IOA is more focused on whole industrial sectors, while LCA is geared towards end-user products, taken as functional units; finally, primary materials (such as metals), waste and emissions are more easily characterized through MFA. Methodological choices are driven by the type of question one needs to address, data availability and collection method and the spatial scales under consideration. Indeed, data can be used in two different ways: bottom-up or top-down. The bottom-up data is more precise, but in general precludes comprehensiveness; on the contrary, the top-down data is by nature more comprehensive, but is not suited for a detailed, fine-scale analysis of the results.

STEEP is pursuing its research program on this theme with three major goals: 1) Creating a comprehensive database enabling pressure analyses; 2) Developing methodologies and models resolving scaling issues, and developing algorithms allowing us to rigorously and automatically obtain adequate assessments; 3) Providing a synthetic analysis of environmental pressures associated to the major material flows, at various geographic levels (employment catchment area, département and région, for France), with the explicit aim of incorporating this type of information in the public decision process on environmental issues, via specifically designed decision-help procedures.

4.3. Urban economy and land use/land cover changes: assessment of spatial distributions of the pressures

The preceding section was focused on territorial metabolism, in particular on the analysis of supply chains. Here territories are examined with a more prominent emphasis on their spatial dimension, with attention to: the spatial distribution of local pressures previously identified (from a land use point of view), and the modeling of future land use and activity location (from an economic point of view). These two questions correspond to very different modeling strategies: the first one is more statistical in nature, extrapolating future land use from past evolution combined with global territory scenarios; the other one has a more fundamental flavor and focuses on an understanding of the processes driving urbanization. For this, we focus more precisely on the question of household and businesses choices of localization, as well as on spatial fluxes within the territory (transportation of goods and persons). The critical point here is to understand and manage urban sprawl and its environmental effects (GHG emission, loss of arable land, ecosystem fragmentation, and so on).

4.3.1. Land Use/Land Cover Change models (LUCC)

LUCC models are mostly used in environmental sciences, e.g. to evaluate the impact of climate change on agriculture, but they can also be used to analyze urban sprawl. There is a variety of models, static or dynamic, grid- or agent- based, local or global, etc., and with varying degrees of sophistication concerning spatio-temporal analysis or decision structures incorporated in the model.

The models of interest here are statistical in nature but spatially explicit. Following decades of development, they are robust, versatile and mature. In principle, agent-models have a larger potential for representing decision processes, but in practice this advantage results in a loss of universality of the models. Among the most well-known and most mature models, one can mention the CLUE family of models, DINAMIC, or LCM (Land Change Modeler). These models are well described in the literature, and will only be briefly presented here.

These models analyze change in land use in a statistical way; they are structured around three different modules:

- The first module determines the probability of change of pixels of the territory (pixels are typically tens to hundreds of meters in size).
- The second module defines the global changes between the various land uses of interest per time step (usually, a few years), based on global scenarios of evolution of the territory under study. These first two modules are independent of one another.
- The last module distributes changes of land use in an explicit manner, pixel per pixel, at each time step, on the basis of the information provided by the first two modules.

Probabilities of change are calibrated on past evolution, from the differences between two past maps of land use in the more favorable cases, or from a single map otherwise (under the assumption that the logic of occupation changes is the same as the logic of land use at this single date). Such changes are then characterized in a statistical way with the help of modeling variables identified by the modeler as having potential explaining or structuring power (typically, a few to a dozen variables are used for one type of land use change). For example, in the case of urban sprawl, typical explaining factors are the distance to existing urbanized zones or distances to roads and other means of transportation, elements of real estate costs, etc. Global scenarios are quantified in terms of global changes in land use over the whole studied area (e.g., how many hectares are

transformed from agricultural to urban uses in a given number of years, how does this evolve over time...); this is done either from academic expert knowledge, or from information provided by local planning agencies. Whenever feasible, models are validated by comparing the model predictions with actual evolution at a later date. Therefore, such models need from one to three land use maps at different dates for calibration and validation purposes (the larger the number of maps, the more robust and accurate the model). A large array of statistical tools is available in the literature to perform the calibration and validation of the model.

The horizon of projections of such models is limited in time, typically 20-30 years, due to the inherent uncertainty in such models, although they are occasionally used on longer time-scales. Climate change constraints are included, when needed, through scenarios, as it is not in the scope of such models to incorporate ecological processes that may translate climate change constraints into land cover change dynamics. Note that on such short time-scales, climate change is not dominated by the mean climate evolution but by decade variations which average out on longer time-scales and are not modeled in the global climate models used e.g. for IPCC projections for the end of the century; as a consequence, the various IPCC climate scenarios cannot be distinguished on such a short time horizon.

With regard to LUCC, the STEEP team has been involved for five years in the ESNET project whose funding came to a close in July of 2017, but the scientific production of the project is still underway. This project bears on the characterization of local Ecosystem Services networks; the project has been coordinated by LECA (Laboratoire d'Ecologie Alpine), in collaboration with a number of other research laboratories (most notably, IRSTEA Grenoble, besides our team), and in close interaction with a panel of local stakeholders; the scale of interest is typically a landscape (in the ecologic/geographic sense, i.e., a zone a few kilometers to a few tens of kilometers wide). The project aims at developing a generic modelling framework of ecosystem services, and studying their behavior under various scenarios of coupled urban/environment evolution, at the 2030/2040 horizon, under constraints of climate change. The contribution of the STEEP team is centered on the Land Use/Land Cover Change (LUCC) model that is one of the major building blocks of the whole project modelling effort, with the help of an ESNET funded post-doctoral researcher. In the process, areas of conceptual and methodological improvements of statistical LUCC models have been identified; implementing these improvements will be useful for the LUCC community at large, independently of the ESNET project needs.

4.3.2. Models for Land-Use and Transportation Interactions (LUTI)

Urban transport systems are intricately linked to urban structure and activities, i.e., to land use. Urbanization generally implies an increased travel demand. Cities have traditionally met this additional demand by extending transportation supply, through new highways and transit lines. In turn, an improvement of the accessibility of ever-farther land leads to an expansion of urban development, resulting in a significant feedback loop between transportation infrastructure and land use, one of the main causes of urban sprawl. Transportation models allow us to address questions generally limited to the impacts of new infrastructures, tolls and other legislation on traffic regulation [0], on user behavior [0], or on the environment [0]. LUTI models (Land-Use and Transport Integrated models) can answer a much broader spectrum of issues. For example, they allow us to understand how the localization of households and of economic activities (which generate transportation demand) adapt to changes of transportation supply. They also allow us to assess the impacts of such changes on the increase in real estate value, or more generally on their effects on the economic development of a specific sector or neighborhood. An economic vision interprets all these interactions in terms of equilibrium between demand and supply. Modelling the localization of households and employments (companies) relies on capturing the way stakeholders arbitrate between accessibility, real estate prices, and attractiveness of different areas.

State of the art and operability of LUTI models. The first model that proved able to analyze the interactions between transport and urbanization was developed by Lowry. Since then theories and models have become increasingly complex over time. They can be classified according to different criteria. A first classification

[0] Congestion, cost and time spent for the transport, etc.
[0] Changes in modality choice.
[0] CO2 emissions, air pollution, noise nuisance, etc.

retraces the historic path of these theories and models. They can be associated with one or several of the approaches underlying all present theories: economic base theory and gravity models, Input/Output models and theory of urban rent, and micro-simulations. A second possibility consists in classifying the models according to their aims and means. Significant scientific progress has been made over the last thirty years. Nevertheless, modelling tools remain largely restricted to the academic world. Today, only seven models have at least had one recent application outside academia or are commercialized or potentially marketable, in spite of the important needs expressed by the urban planning agencies: Cube Land, DELTA, MARS, OPUS/UrbanSim, PECAS, TRANUS and Pirandello.

To guide their choice of a modelling framework, users can rely on various criteria such as the strength of the theoretical framework, the quality and the diversity of the available documentation, the accessibility of the models (is the model freely available? is the code open source? is the software regularly updated and compatible with the recent operating systems?), the functionality and friendliness of user interfaces (existence of graphic user interface, possibility of interfacing with Geographic Information Systems), existence of technical assistance, volume and availability of the data required to implement the model, etc. For example, among the seven models mentioned above, only two are open source and mature enough to meet professional standards: TRANUS and UrbanSim [0]. These two models are very different but particularly representative of the main current philosophies and trends in this scientific domain. Their comparison is informative.

STEEP implication in LUTI modelling. As yet, very few local planning authorities make use of these strategic models, mostly because they are difficult to calibrate and validate. Systematic improvement on these two critical steps would clearly increase the level of confidence in their results; these limitations hinder their dissemination in local agencies. One of the major goals of STEEP is therefore to meet the need for better calibration and validation strategies and algorithms. This research agenda lies at the core of our project CITiES (ANR Modèles Numériques) that ended in 2017 with the PhD defense of Thomas Capelle . This work is being partly pursued in the QAMECS project.

As for LUTI modeling, we have been using the TRANUS model since the creation of our team. In this framework we work in close collaboration with AURG [0], the local urban planning agency of Grenoble (*Agence d'Urbanisme de la Région Grenobloise*) in order to better understand and to improve the relevance of these tools for such territorial agencies.

[0] http://www.urbansim.org
[0] http://www.aurg.org/

TONUS Project-Team

4. Application Domains

4.1. Controlled fusion and ITER

The search for alternative energy sources is a major issue for the future. Among others, controlled thermonuclear fusion in a hot hydrogen plasma is a promising possibility. The principle is to confine the plasma in a toroidal chamber, called a tokamak, and to attain the necessary temperatures to sustain nuclear fusion reactions.The International Thermonuclear Experimental Reactor (ITER) is a tokamak being constructed in Cadarache, France. This was the result of a joint decision by an international consortium including the European Union, Canada, USA, Japan, Russia, South Korea, India and China. ITER is a huge project. As of today, the budget is estimated at 20 billion euros. The first plasma shot is planned for 2020 and the first deuterium-tritium operation for 2027. Many technical and conceptual difficulties have to be overcome before the actual exploitation of fusion energy. Consequently, much research has been carried out around magnetically confined fusion. Amongt hese studies, it is important to carry out computer simulations of the burning plasma. Thus, mathematicians and computer scientists are also needed in the design of ITER. The reliability and the precision of numerical simulations allow a better understanding of the physical phenomena and thus would lead to better designs. TONUS's main involvement is in such research. The required temperatures to attain fusion are very high, of the order of a hundred million degrees. Thus it is imperative to prevent the plasma from touching the tokamak inner walls. This confinement is obtained thanks to intense magnetic fields. The magnetic field is created by poloidal coils, which generate the toroidal component of the field. The toroidal plasma current also induces a poloidal component of the magnetic field that twists the magnetic field lines. The twisting is very important or the stability of the plasma. The idea goes back to research by Tamm and Sakharov, two Russian physicists, in the 50's. Other devices are essential for the proper operation of the tokamak: divertor for collecting the escaping particles, microwave heating for reaching higher temperatures, fuel injector for sustaining the fusionr eactions, toroidal coils for controlling instabilities, etc.

4.2. Other applications

The software and numerical methods that we develop can also be applied to other fields of physics or of engineering.

- For instance, we have a collaboration with the company AxesSim in Strasbourg for the development of efficient Discontinuous Galerkin (DG) solvers on hybrid computers (CPU/GPU). The applications are electro-magnetic simulations for the conception of antennas, electronic devices or aircraft electromagnetic compatibility.
- The acoustic conception of large rooms requires huge numerical simulations. It is not always possible to solve the full wave equation and many reduced acoustic models have been developed. A popular model consists in considering "acoustic" particles moving at the speed of sound. The resulting Partial Differential Equation (PDE) is very similar to the Vlasov equation. The same modelling is used in radiation theory. We have started to work on the reduction of the acoustic particles model and realized that our reduction approach perfectly applies to this situation. A PhD with CEREMA (Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement) has started in October 2015 (PhD of Pierre Gerhard). The objective is to investigate the model reduction and to implement the resulting acoustic model in our DG solver.
- In September 2017, we started a collaboration with EDF Chatou (PhD of Lucie Quibel) on the modelling of multiphase fluids with complex equations of state. The goal is to simulate the high temperature liquid-vapor flow occurring in a nuclear plant. Among others, we will apply our recent kinetic method for designing efficient implicit schemes for this kind of flows.

BIOCORE Project-Team

4. Application Domains

4.1. Bioenergy

Finding sources of renewable energy is a key challenge for our society. We contribute to this topic through two main domains for which a strong and acknowledged expertise has been acquired over the years. First, we consider anaerobic digesters, the field of expertise of the members of the team at the Laboratory of Enviromental Biotechnology (LBE), for the production of methane and/or biohydrogen from organic wastes. The main difficulty is to make these processes more reliable and exploit more efficiently the produced biogas by regulating both its quality and quantity despite high variability in the influent wastes. One of the specific applications that needs to be tackled is the production of biogas in a plant when the incoming organic waste results from the mixing of a finite number of substrates. The development of control laws that optimize the input mix of the substrates as a function of the actual state of the system is a key challenge for the viability of this industry.

The second topic consists in growing microalgae, the field of expertise of the members of the team at the Oceanographic Laboratory of Villefranche-sur-Mer (LOV), to produce biofuel. These microorganisms can synthesize lipids with a much higher productivity than terrestrial oleaginous species. The difficulty is to better understand the involved processes, which are mainly transient, to stimulate and optimize them on the basis of modeling and control strategies. Predicting and optimizing the productivity reached by these promising systems in conditions where light received by each cell is strongly related to hydrodynamics, is a crucial challenge.

Finally, for the energy balance of the process, it is important to couple microalgae and anaerobic digestion to optimize the solar energy that can be recovered from microalgae, as was explored within the ANR Symbiose project (2009-2012) [3].

4.2. CO_2 fixation and fluxes

Phytoplanktonic species, which assimilate CO_2 during photosynthesis, have received a lot of attention in the last years. Microalgal based processes have been developed in order to mitigate industrial CO_2. As for biofuel productions, many problems arise when dealing with microalgae which are more complex than bacteria or yeasts. Several models have been developed within our team to predict the CO_2 uptake in conditions of variable light and nitrogen availability. The first modeling challenge in that context consists in taking temperature effects and light gradient into account.

The second challenge consists in exploiting the microalgal bioreactors which have been developed in the framework of the quantification of carbon fluxes between ocean and atmospheres. The SEMPO platform (simulator of variable environment computer controlled), developed within the LOV team, has been designed to reproduce natural conditions that can take place in the sea and to accurately measure the cells behavior. This platform, for which our team has developed models and control methods over the years, is an original and unique tool to develop relevant models which stay valid in dynamic conditions. It is worth noting that a better knowledge of the photosynthetic mechanisms and improved photosynthesis models will benefit both thematics: CO_2 mitigation and carbon fluxes predictions in the sea.

4.3. Biological control for plants and micro-plants production systems

This research concentrates on the protection of cultures of photosynthetic organisms against their pests or their competitors. The cultures we study are crop and micro-algae productions. In both cases, the devices are more or less open to the outside, depending on the application (greenhouse/field, photobioreactor/raceway), so that they may give access to harmful pathogens and invading species. We opt for protecting the culture through the use of biocontrol in a broad sense.

In crop production, biocontrol is indeed a very promising alternative to reduce pesticide use: it helps protecting the environment, as well as the health of consumers and producers; it limits the development of resistance (in comparison to chemicals). The use of biocontrol agents, which are, generically, natural enemies (predators, parasitoids or pathogens) of crop pests [89], is however not widespread yet because it often lacks efficiency in real-life crop production systems (while its efficiency in the laboratory is much higher) and can fail to be economically competitive. Resistant crops are also used instead of pesticides to control pests and pathogens, but the latter eventually more or less rapidly overcome the resistance, so these crops need to be replaced by new resistant crops. As resistant genes are a potentially limited resource, a challenge is to ensure the durability of crop resistance. Our objective is to propose models that would help to explain which factors are locks that prevent the smooth transition from the laboratory to the agricultural crop, as well as develop new methods for the optimal deployment of the pests natural enemies and of crop resistance.

Microalgae production is faced with exactly the same problems since predators of the produced microalgae (e.g. zooplankton) or simply other species of microalgae can invade the photobioreactors and outcompete or eradicate the one that we wish to produce. Methods need therefore to be proposed for fighting the invading species; this could be done by introducing predators of the pest and so keeping it under control, or by controling the conditions of culture in order to reduce the possibility of invasion; the design of such methods could greatly take advantage of our knowledge developed in crop protection since the problems and models are related.

4.4. Biological depollution

These works will be carried out with the LBE , mainly on anaerobic treatment plants. This process, despite its strong advantages (methane production and reduced sludge production) can have several locally stable equilibria. In this sense, proposing reliable strategies to stabilize and optimise this process is a key issue. Because of the recent (re)development of anaerobic digestion, it is crucial to propose validated supervision algorithms for this technology. A problem of growing importance is to take benefit of various waste sources in order to adapt the substrate quality to the bacterial biomass activity and finally optimize the process. This generates new research topics for designing strategies to manage the fluxes of the various substrate sources meeting at the same time the depollution norms and providing a biogas of constant quality. In the past years, we have developed models of increasing complexity. However there is a key step that must be considered in the future: how to integrate the knowledge of the metabolisms in such models which represent the evolution of several hundreds bacterial species? How to improve the models integrating this two dimensional levels of complexity? With this perspective, we wish to better represent the competition between the bacterial species, and drive this competition in order to maintain, in the process, the species with the highest depollution capability. This approach, initiated in [92] must be extended from a theoretical point of view and validated experimentally.

CARMEN Project-Team

4. Application Domains

4.1. Scientific context: the LIRYC

The University Hospital of Bordeaux (*CHU de Bordeaux*) is equipped with a specialized cardiology hospital, the *Hôpital Cardiologique du Haut-Lévêque*, where the group of Professor Michel Haïssaguerre has established itself as a global leader in the field of cardiac electrophysiology [63], [61], [48]. Their discoveries in the area of atrial fibrillation and sudden cardiac death syndromes are widely acclaimed, and the group is a national and international referral center for treatment of cardiac arrhythmia. Thus the group also sees large numbers of patients with rare cardiac diseases.

In 2011 the group has won the competition for a 40 million euro *Investissements d'Avenir* grant for the establishment of IHU Liryc, an institute that combines clinical, experimental, and numerical research in the area of cardiac arrhythmia (http://ihu-liryc.fr). The institute works in all areas of modern cardiac electrophysiology: atrial arrhythmias, sudden death due to ventricular fibrillation, heart failure related to ventricular dyssynchrony, and metabolic disorders. It is recognized as one of the most important centers worldwide in this area.

The Carmen team was founded to partner with IHU Liryc. We bring applied mathematics and scientific computing closer to experimental and clinical cardiac electrophysiology. In collaboration with experimental and clinical researchers at Liry we work to enhance fundamental knowledge of the normal and abnormal cardiac electrical activity and of the patterns of the electrocardiogram, and we develop new simulation tools for training, biological, and clinical applications.

4.2. Basic experimental electrophysiology

Our modeling is carried out in coordination with the experimental teams from IHU Liryc. It help to write new concepts concerning the multiscale organisation of the cardiac action potentials that will serve our understanding in many electrical pathologies. For example, we model the structural heterogeneities at the cellular scale [50], and at an intermediate scale between the cellular and tissue scales.

At the atrial level, we apply our models to understand the mechanisms of complex arrythmias and the relation with the heterogeneities at the insertion of the pulmonary veins. We will model the heterogeneities specific to the atria, like fibrosis or fatty infiltration [69] [60]. These heterogeneities ara thought to play a major role in the development of atrial fibrillation.

At the ventricular level, we focus on (1) modeling the complex coupling between the Purkinje network and the ventricles, which is supposed to play a major role in sudden cardiac death, and (2) modeling the heteogeneities related to the complex organization and disorganization of the myocytes and fibroblasts, which is important in the study of infarct scars for instance.

4.3. Clinical electrophysiology

Treatment of cardiac arrhythmia is possible by pharmacological means, by implantation of pacemakers and defibrillators, and by curative ablation of diseased tissue by local heating or freezing. In particular the ablative therapies create challenges that can be addressed by numerical means. Cardiologists would like to know, preferably by noninvasive means, where an arrhythmia originates and by what mechanism it is sustained.

We address this issue in the first place using inverse models, which attempt to estimate the cardiac activity from a (high-density) electrocardiogram. A new project aims at performing this estimation on-site in the catheterization laboratory and presenting the results, together with the cardiac anatomy, on the screen that the cardiologist uses to monitor the catheter positions [66].

An important prerequisite for this kind of interventions and for inverse modeling is the creation of anatomical models from imaging data. The Carmen team contributes to better and more efficient segmentation and meshing through the IDAM project.

COMMEDIA Project-Team

4. Application Domains

4.1. Cardiovascular hemodynamics

The heart is a double pump whose purpose is to deliver blood to the tissue and organs of the body. This function is made possible through the opening and closing of the heart valves. Cardiac diseases generally manifest by affecting the pumping function of the heart. Numerical simulations of cardiac hemodynamics, in normal and pathological conditions, are recognized as a tool of paramount importance for improving the understanding, diagnosis and treatment of cardiac pathologies, and also for the development of implantable devices (see, e.g., [63], [46]). As an example, we can mention the case of cardiac mitral valve regurgitation, one of the most common heart valve diseases. For this pathology, clinical data are known to be insufficient for determining the optimal timing for surgery, the best surgical strategy and the long-term outcome of a surgical repair. Contrary to imaging techniques, numerical simulations provide local information, such as pressure and stresses, which are of fundamental importance for the prediction of the mechanical behavior of native valves and of implantable devices.

4.2. Respiratory flows

Respiration involves the transport of air through the airways from the mouth to the alveoli of the lungs. These units where diffusion of oxygen and carbon dioxide take place, are surrounded by a viscoelastic medium (the parenchyma) consisting of blood vessels and collagen fibers. Air flows due to the displacement of the diaphragm, which drives the pulmonary parenchyma. Accidental inhalations of foreign bodies or pathologies such as asthma, emphysema and fibrosis might prevent the lung of fulfilling its function. Therapies mostly use aerosols (set of small particles, solid or liquid), which must reach the specific areas of the lung targeted for treatment. Understanding the airflow mechanisms within the respiratory network is a fundamental ingredient for predicting the particles motion and their deposition (see, e.g., [44]). Moreover, understanding of the gas diffusion in the lung is also of major importance since the main fonction of this organ is to deliver oxygen to the blood.

4.3. Safety pharmacology

The problem of safety pharmacology can be summarized as follows: given a molecule which is a candidate to become a drug, is its use dangerous due to side effects? Among all the different problems to be addressed, one of the most relevant questions in pharmacology is cardio-toxicity (see [58]). More precisely, the objective is to determine whether or not a molecule alters in a significant way the normal functioning of the cardiac cells. To answer these questions, the CiPA initiative promotes the introduction of novel techniques and their standardisation (see [50]). One of the proposed tests of the CiPA panel is to measure the the electrical activity using Micro-Electrodes Array: these are microchips that record the electrical activity of an ensemble of cells. The task is to infer the impact of a drug on the ionic currents of each cell based on the electrical signal measured (electrograms) and, in perspective, to be able to assess whether a molecule can induce arrhythmia (see [49]).

DRACULA Project-Team (section vide)

M3DISIM Project-Team (section vide)

MAMBA Project-Team

4. Application Domains

4.1. Introduction

The team has two main application-driven research axes. Applicative axis 1 focuses on cancer, an application on which almost all team members work, with various approaches. A main focus of the team is to study cancer as a Darwinian evolutionary phenomenon in phenotype-structured cell populations. Optimal control methods take into account the two main pitfalls of clinical cancer therapeutics, namely unwanted toxic side effects in healthy cell populations and drug resistance in cancer cell populations. Other studies concern telomere shortening, and multi-scale models. Applicative axis 2 is devoted to growth, evolution and regeneration in populations and tissues. It involves protein aggregation and fragmentation models for neurodegenerative diseases (prion, Alzheimer), organ modeling, mainly of the liver, its damages induced by toxic molecules, and its regeneration after toxic insult. Newcomers in this applicative field are epidemiological modeling of propagation of insect vector-borne diseases by reaction-diffusion equations and of their optimal control, bacterial growth and wound healing.

4.2. Applicative axis 1: Focus on cancer

Personnel
Luis Almeida, Jean Clairambault, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil.

Project-team positioning

The MAMBA team designs and analyses mathematical models of tumor growth and therapy, at the cell population level, using agent-based or partial differential equations, with special interest in methodologies for therapeutic optimization using combined anticancer drug treatments. Rather than, or not only, modeling the effect of drugs on molecular targets, we represent these effects by their functional consequences on the fate of healthy and cancer cell populations: proliferation (velocity of the cell division cycle, decreasing it, e.g., by antagonizing growth factor receptors), apoptosis, cell death or senescence. Our goal in doing this is to circumvent the two main issues of anticancer therapy in the clinic, namely unwanted toxic side effects in populations of healthy cells and emergence of drug-induced drug resistance in cancer cell populations. This point of view leads us to take into account phenomena of transient and reversible resistance, observed in many cancer cell populations, by designing and analyzing models of cell populations structured in continuous phenotypes, relevant for the description of the behavior of cell populations exposed to drugs: either degree of resistance to a given drug, or potential of resistance to drug-induced stress, proliferation potential, and plasticity. Such modeling options naturally lead us to to take into account in a continuous way (i.e., by continuous-valued phenotype or relevant gene expression) the wide phenotypic heterogeneity of cancer cell populations. They also lead us to adopt the point of view of adaptive dynamics according to which characteristic traits of cell populations evolve with tumor environmental pressure (drugs, cytokines or metabolic conditions, mechanical stress and spatial conditions), in particular from drug sensitivity to resistance. This position is original on the international scene of teams dealing with drug resistance in cancer.

Scientific achievements

Modeling Acute Myeloid Leukemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations
In collaboration with Catherine Bonnet (Inria DISCO, Saclay) and François Delhommeau (St Antoine hospital in Paris), together with DISCO PhD students José Luis Avila Alonso and Walid Djema, this theme has led to common published proceedings of conferences: IFAC, ACC, CDC, MTNS [66], [67], [68], [77], [95], [65]. These works study the stability of the haematopoietic system and its possible restabilization by combinations of anticancer drugs with functional targets on cell populations: proliferation, apoptosis, differentiation.

Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control
We tackle the problem to represent and inhibit - using optimal control algorithms, in collaboration with Emmanuel Trélat, proposed Inria team CAGE - drug-induced drug resistance in cancer cell populations. This theme, presently at the core of our works on cancer modeling with a evolutionary perspective on tumor heterogeneity, is documented in a series of articles [90], [92], [124], [125], [127]. Taking into account the two main pitfalls of cancer therapy, unwanted side effects on healthy cells and evolution towards resistance in cancer cells, it has attracted to our team the interest of several teams of biologists, with whom we have undertaken common collaborative works, funded by laureate answers to national calls (see ITMO Cancer HTE call).

This theme is also at the origin of methodological developments (see Research axis 1). In collaboration with Shensi Shen from Institut Gustave Roussy and Francois Vallette from Université de Nantes, we aim to develop simple non-spatial models to understand the mechanisms of drug resistance acquisition -and loss- in melanoma and glioblastoma. The models are systematically compared with in vitro and in vivo data generated by our collaborators and treated via image processing techniques developed in the team.

Senescence modeling by telomere shortening
In many animals, aging tissues accumulate senescent cells, a process which is beneficial to protect from cancer in the young organism. In collaboration with Teresa Teixeira and Zhou Xu from IBCP, we proposed a mathematical model based on the molecular mechanisms of telomere replication and shortening and fitted it on individual lineages of senescent Saccharomyces cerevisiae cells, in order to decipher the causes of heterogeneity in replicative senescence [79].

Biomechanically mediated growth control of cancer cells
Model simulations indicate that the response of growing cell populations on mechanical stress follows a simple universal functional relationship and is predictable over different cell lines and growth conditions despite the response curves look largely different. We developed a hybrid model strategy in which cells were represented by coarse-grained individual units calibrated in a high resolution cell model and parameterized each model cell by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics. Our model simulation results suggest that the growth response of cell population upon externally applied mechanical stress is the same, as it can be quantitatively predicted using the same growth progression function [123].

Bio-mechanical models of tissue growth
The degenerate Cahn-Hilliard equation is a standard model to describe living tissues. It takes into account cell populations undergoing short-range attraction and long-range repulsion effects. In this framework, we consider the usual Cahn-Hilliard equation with a singular single-well potential and degenerate mobility. These degeneracy and singularity induce numerous difficulties, in particular for its numerical simulation. To overcome these issues, we propose in [hal-02274417] a relaxation system formed of two second order equations which can be solved with standard packages. This system is endowed with an energy and an entropy structure compatible with the limiting equation. Here, we study the theoretical properties of this system; global existence and convergence of the relaxed system to the degenerate Cahn-Hilliard equation. We also study the long-time asymptotics which interest relies on the numerous possible steady states with given mass.

Free boundary multiphase models of tumor growth

Multiphase mechanical models are now commonly used to describe living tissues including tumour growth. The specific model we study here consists of two equations of mixed parabolic and hyperbolic type which extend the standard compressible porous media equation, including cross-reaction terms. We study the incompressible limit, when the pressure becomes stiff, which generates a free boundary problem. We establish the complementarity relation and also a segregation result. Several major mathematical difficulties arise in the two species case which are addressed in [43]. Firstly, the system structure makes comparison principles fail. Secondly, segregation and internal layers limit the regularity available on some quantities to BV. Thirdly, the Aronson-Bénilan estimates cannot be established in our context. We are lead, as it is classical, to add correction terms. This procedure requires technical manipulations based on BV estimates only valid in one space dimension. Another novelty is to establish an L^1 version in place of the standard upper bound.

Philosophy of cancer

The quite natural idea that cancer is a disease of the control of coherent multicellularity, expressed when cohesion of tissues and coherence of (unknown, except maybe for the case of a centralised circadian clock) synchronising signals fail to ensure it, by a regression towards unicellularity, stopping in this "reverse evolution path" at a coarse, incoherent multicellularity state [0] continues to be developed and popularised by Jean Clairambault in seminars and workshops, and published in review articles [13], [45]. This view, and the investigation of the immune system in the design of such coherence of all multicellular organisms [0] is naturally inscribed in a *philosophy of cancer* perspective, and from a mathematical viewpoint, to multicellularity genes - and links between them and unicellularity genes - seen as a *hyperstructure* [0] above structures consisting of the genes of unicellularity, i.e., those that make a single cell a coherent living system, such hyperstructure being failed in cancer; this view is presently under development with colleagues from universities of the Paris region, together with Nils Baas at NTNU, Trondheim, Norway). This perspective, that makes use of category theory as a structuring point of view to apprehend multicellularity and cancer, is also meant to endow us with an innovative methodology to apply *topological data analysis (TDA)* to investigate cancer genome data.

Modelling of TMZ induced drug resistance

Temozolomide (TMZ) is a standard chemotherapy treatment in patients with glioblastoma. Resistance to this drug is correlated to the presence of a specific enzyme, which activity in cancer cells creates a drug-induced cell death resistant phenotype. Understanding the transition of cancer cells to a resistant phenotype is still a topic of research where multiple hypothesis have been studied: From an adaptive process to an inherent resistance to treatment. Moreover it has been recently shown that TMZ treatment has an influence on the spatial structuration of cancer cell aggregates. In the frame of the HTE project MoGlImaging and through the recent hiring of a post-doctoral candidate (Gissell Estrada Rodriguez), we are currently developping a mathematical framework to study and analyse the evolution of a population of glioblastoma cells that are exposed to TMZ. Based on the experimental data generated by our partner team led by F. Valette (Inserm Nantes), we propose a Keller-Segel type model for the formation of spheroid as a result of a chemoattractant produced by cancer cells, and study the influence of a chemotherapeutic agent on the structuration of the cancer cell aggregates. By confronting the model results to experimental data, different modelling choices are currently explored to identify which key mechanisms could be responsible for the apparition of drug resistance in glioblastoma.

Collaborations

- AML modelling: **Catherine Bonnet**, DISCO Inria team, Saclay, and **François Delhommeau**, INSERM St Antoine (also collaborator in the INSERM HTE laureate project EcoAML, see below).
- INSERM HTE laureate project MoGlImaging, headed by E. Moyal (Toulouse): **François Vallette**, CRCNA and INSERM Nantes
- INSERM HTE laureate project EcoAML, headed by **François Delhommeau**, INSERM St Antoine: François Delhommeau, Thierry Jaffredo (IBPS), Delphine Salort (LCQB-IBPS)
- Adaptive dynamics to model drug resistance and optimal control to circumvent it:

[0] Metazoa 1.0, as theorised by PCW Davies and CH Lineweaver in their article "Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors", Physical Biology 2011, that popularised the so-called atavistic hypothesis of cancer

[0] this latter point partly, however nicely, developed in Thomas Pradeu's book "The limits of the self", OUP 2012

[0] See on this point, e.g., Nils Baas: "On the philosophy of higher structures", Int. J. General Systems 2019

Alexandre Escargueil, Michèle Sabbah (1 PhD thesis in common), St Antoine Hospital, Paris

Emmanuel Trélat (1 PhD thesis in common) at Inria team CAGE and Laboratoire Jacques-Louis Lions at Sorbonne Université.

Frédéric Thomas at CREEC, Montpellier.

Tommaso Lorenzi (Univ. of St Andrews).

- Telomere shortening: **Teresa Teixeira** and **Zhou Xu** (IBCP, Paris), **Philippe Robert** (Inria RAP).
- TRAIL treatment: **Gregory Batt,** Inria Saclay and Inst. Pasteur (France)
- Biomechanical control of cancer cells: **Pierre Nassoy**, Bioimaging and Optofluidics Group, LP2N – UMR 5298. IOGS, CNRS & University of Bordeaux

4.3. Applicative axis 2: Growth, evolution and regeneration in populations and tissues

Luis Almeida, Pierre-Alexandre Bliman, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil, Philippe Robert

Project-team positioning

The applications in this category span very different subjects from amyloid diseases, dengue fever, wound healing, liver regeneration and toxicity, up to bacterial growth and development of organisms. As the applications, the methods span a wide range. Those concerning identification of models and parameters with regard to data have partially been outlined in axis 3. Focus in this axis is on the model contribution to the biologically and/or medically relevant insights and aspects.
Liver-related modelling is partially performed within the Inria team MIMESIS (Strasbourg) with the focus on real-time, patient-specific biomechanical liver models to guide surgery and surgeons. Internationally, spatial temporal liver related models are developed in Fraunhofer MEVIS (Bremen), by T. Ricken (TU Dortmund), and P. Segers group (Leuven).

Different from these, Mamba has a strong focus on spatial-temporal modeling on the histological scale, integration of molecular processes in each individual cell, and single-cell (agent) based models [102]. Works by Schliess [149], [107] have been highlighted in editorials.

Mathematical modeling of protein aggregation is a relatively recent domain, only a few other groups have emerged yet; among them we can cite the Inria team Dracula, with whom we are in close contact, and e.g., the work by Jean-Michel Coron (Sorbonne Université) and Monique Chyba (Hawaii, USA) in control, and Suzanne Sindi (USA) for the modeling of the yeast prion. We have interactions with all these groups and organized a workshop in June 2017, gathering both the biophysics and applied mathematics communities.

Scientific achievements

Amyloid disease
Application to protein aggregation in amyloid diseases is a long-standing interest of Mamba, dating back to 2010 [85], and developed through the collaboration with Human Rezaei's team at Inra. More recently, with Wei-Feng Xue in Canterbury, we investigated the intrinsic variability among identical experiments of nucleation [98], [106], Sarah Eugène's Ph.D subject (co-supervised by Philippe Robert) [105].

In collaboration with Tom Banks first [70], [69] and then Philippe Moireau, we developed quantitative comparisons between model and data. Through data assimilation and statistical methods [63], we proposed new models and mechanisms.

Biological control of arboviroses
Sterile Insect Technique (SIT) [104] is a biological control method relying on massive releases of sterile male insects into the wild. The latter compete with wild males to mate with the females, and induce no offspring to the latter, thus reducing the next generation's population. This can result in a progressive reduction, or even disparition, of the target population.

A related technique is based on the infection by *Wolbachia* [111]. This symbiotic bacterium is maternally transmitted from infected females to their offspring, but induces *cytoplasmic incompatibility* [151], [80]: mating between infected males and uninfected females gives no offspring. Releases of *Wolbachia* infected males alone is thus comparable to classical SIT.

On the other hand, releasing both infected males and females in sufficient quantity may result in infection of the wild population. This gives rise to an interesting new control principle, as *Wolbachia* has been shown to severely reduce the insect vectorial ability to transmit dengue, zika or chikungunya, indirectly by lifespan and fertility reduction, and directly by reducing the ability of the viruses to proliferate within the organism [130].

We proposed new insights on the practical and theoretical issues raised by the implementation of the previous methods. Concerning the SIT, we obtained control synthesis results through impulsive periodic release of controlled amplitude [10], and through optimal control approach [42]. Concerning Wolbachia technique, we investigated general control principles [39] capable of spreading the infection.

Wound healing 1: epithelial tissues
We studied cell motion in epithelial gap closure, a form of collective cell migration that is a very widespread phenomenon both during development and adult life - it is essential for both the formation and for the maintenance of epithelial layers. Due to their importance, *in vivo* wound healing and morphogenetic movements involving closure of holes in epithelia have been the object of many studies. In our works [0] we considered wound healing and epithelial gap closure in both in vivo (in particular drosophila pupa) and in vitro (MDCK cell and human keratinocytes). We found some similarities in the geometry dependence of the wound closure strategies between these two situations, indicating the existence of conserved mechanisms that should be widespread across living beings. We are concentrating on the study of actin cable formation.

Wound healing 2: adipose tissues
After injury, if regeneration can be observed in hydra, planaria and some vertebrates, regeneration is rare in mammals and particularly in humans. In this research axis, we investigated the mechanisms by which biological tissues recover after injury. We explored this question on adipose tissue, using the mathematical framework recently developed in [144]. Our assumption is that simple mechanical cues between the Extra-Cellular Matrix (ECM) and differentiated cells can explain adipose tissue morphogenesis and that regeneration requires after injury the same mechanisms. We validated this hypothesis by means of a two-dimensional Individual Based Model (IBM) of interacting adipocytes and ECM fiber elements [145]. The model successfully generated regeneration or scar formation as functions of few key parameters, and seemed to indicate that the fate of injury outcome could be mainly due to ECM rigidity.

Following these encouraging results, the team is currently taking a step further in the model validation and confrontation to experimental data. The first direction concerns the development of a 3D framework to validate the mechanisms observed in 2D. In collaboration with S. Merino-Aceituno from the University of Vienna, efforts are being made in the development of a complete synthetic tissue model coupling the ECM and cell modelling with a vascularization model.A PhD project has been started to implement the coupled models and reduce the CPU time with the long-term goal to develop a usable software which would serve to investigate the role of different mechanisms in tissue development (not restricted to adipose tissues). Finally, further developments in collaboration with Imperial College London aim at pursuing the derivation of macroscopic PDE models from the agent-based formalisms.

Mathematical modelling of axolotl regeneration
Tissue response after injury/amputation induces one or two alternatives: scar formation versus regeneration (complete recovery of tissue shape and functions). In most mammals, regeneration is considered largely impaired for the benefit of a fibrotic scar after injury automatically associated with dysfunctions, but complete regeneration has been largely described and investigated in animal models such as zebra fish, salamander, or axolotl. Despite several processes regulating regeneration have been identified at different scales -from diffusing molecules and cellular gene expression patterns up to tissue mechanics-, how these mechanisms individually or collectively play a role in the regulation of regenerative processes remains poorly understood.

[0] ravasio:hal-01245750, vedula:hal-01298859

In order to give insights into the mechanisms of tissue regeneration, Valeria Caliaro started an Inria PhD project in october 2019, in collaboration with Osvaldo Chara, internationally recognized group leader of SysBio in Argentina. This project focuses on the role of cell proliferation in space and time along the two first phases of regeneration after injury: (i) initiation of a regeneration response, (ii) tissue patterning during regenerate growth. The first part of the project aims at building an agent-based model featuring few key mechanisms regulating cell proliferation after injury. The model construction is based on recent works where the authors developed a mathematical model given by ordinary differential equations (ODEs)[2] and a mathematical framework in 1D [3] showing that acceleration of the cell cycle is the major driver of regenerative spinal cord outgrowth in axolotls. Building on both mathematical models and introducing heuristic rules which rely on Prof O. Chara expertise, we propose a 2D-ABM using methodologies borrowed from socio-dynamics and collective behavior studies (based on many interacting agent systems). While the focus is made on proliferation-based mechanisms, other mechanisms responsible for collective behavior such as volume exclusion, diffusion or aggregation will be tested and compared with experimental data. The resulting model will provide a synthetic tissue model which will serve to investigate regeneration in cellular systems, focusing on cell proliferation properties. The second part of the PhD will be devoted to the derivation of continuous models from the agent-based formalism. This will provide a large scale 'synthetic tissue' model to explore the role of large scale effects in general tissue models. Model validation and calibration will be ensured by quantitative comparison with biological data already present in the literature and generated by the SysBio group of O. Chara, particularly the representative images of regenerative spinal cords after tail amputation. By varying the model parameters and observing the resulting alteration of the spinal cord size and architecture as a consequence of these variations, it will be possible to provide 'in silico' setting experiments to guide and plan future in vivo or ex vivo experiments. Altogether, the project is expected to provide a mechanistic understanding of the cellular mechanisms driving spinal cord regeneration, and to identify how spatial structuration can influence cell differentiation and growth.

[1] Rodrigo Albors A, Tazaki A, Rost F, Nowoshilow S, Chara O & Tanaka EM. 2015. Planar cell polarity-mediated induction of neural stem cell expansion during axolotl spinal cord regeneration. Elife. 4:e10230. [2] Rost F, Rodrigo Albors A, Mazurov V, Brusch L, Deutsch A, Tanaka EM & Chara O. 2016. Accelerated cell divisions drive the outgrowth of the regenerating spinal cord in axolotls. Elife. 5. pii: e20357. [3] Cura Costa E, Rodrigo Albors A, Tanaka EM & Chara O. Spatiotemporal distribution of cell proliferation rules spinal cord regeneration in the axolotl. MS in preparation.

Quantitative cell-based model predicts mechanical stress response of growing tumor spheroids

Model simulations indicate that the response of growing cell populations on mechanical stress follows the same functional relationship and is predictable over different cell lines and growth conditions despite experimental response curves look largely different. We developed a hybrid model strategy in which cells are represented by coarse-grained individual units calibrated with a high resolution cell model and parameterized by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics from the growth kinetics in absence of external stress. Our model simulation results suggest a generic, even quantitatively same, growth response of cell populations upon externally applied mechanical stress, as it can be quantitatively predicted using the same growth progression function ([52])

Modeling of morphogen diffusion in *Drosophila* oogenesis

In collaboration with a team of developmental biologists of Rutgers University (Camden, New Jersey), we have built a model for the diffusion of the Gurken morphogen during *Drosophila* oogenesis, taking into account a wide variety of biological mechanisms such as diffusion of the morphogen, reactions of components of the EGFR signaling pathway, movement of the source of morphogen, shift of the overlying follicle cells and growth of the egg chamber. This model, together with a complete numerical code developed in Matlab,

provides a tool to understand how each mechanism influences the signal distribution. The overall aim of the project is to use this tool to guide future experiments, and to understand what mechanisms contribute to the different distributions of signal among species.

Bacterial population growth
We exploited all the methods developed to estimate the division rate of a population (see axis 3) to address a seminal question of biology: is it a size-sensing or a timing mechanism which triggers bacterial growth? In [148], we showed that a sizer model is robust and fits the data well. Several studies from other groups came at the same time, showing a renewed interest on a question dated back to Jacques Monod's PhD thesis (1941). Of special interest is the "adder" model, for which we are currently developing new estimation methods [46].

A quantitative high resolution computational mechanics cell model for growing and regenerating tissues
Mathematical models are increasingly designed to guide experiments in biology, biotechnology, as well as to assist in medical decision making. They are in particular important to understand emergent collective cell behavior. For this purpose, the models, despite still abstractions of reality, need to be quantitative in all aspects relevant for the question of interest. Considered was as showcase example the regeneration of liver after drug-induced depletion of hepatocytes, in which the surviving and dividing hepatocytes must squeeze in between the blood vessels of a network to refill the emerged lesions. Here, the cells' response to mechanical stress might significantly impact the regeneration process. We present a 3D high-resolution cell-based model integrating information from measurements in order to obtain a refined and quantitative understanding of the impact of cell-biomechanical effects on the closure of drug-induced lesions in liver. Our model represents each cell individually and is constructed by a discrete, physically scalable network of viscoelastic elements, capable of mimicking realistic cell deformation and supplying information at subcellular scales. The cells have the capability to migrate, grow, and divide, and the nature and parameters of their mechanical elements can be inferred from comparisons with optical stretcher experiments. Due to triangulation of the cell surface, interactions of cells with arbitrarily shaped (triangulated) structures such as blood vessels can be captured naturally. Comparing our simulations with those of so-called center-based models, in which cells have a largely rigid shape and forces are exerted between cell centers, we find that the migration forces a cell needs to exert on its environment to close a tissue lesion, is much smaller than predicted by center-based models. To stress generality of the approach, the liver simulations were complemented by monolayer and multicellular spheroid growth simulations. In summary, our model can give quantitative insight in many tissue organization processes, permits hypothesis testing in silico, and guide experiments in situations in which cell mechanics is considered important [123].

Liver regeneration and disease: towards a full virtual liver model at histological scale
In our work towards a full virtual liver model at histological level, a number of steps were performed. The models under points (1)-(4) focus on either a single or a few liver lobules. A liver lobule is the smallest repetitive functional and anatomical building block of liver, while (5) addresses a much larger organisational building block of the liver, a liver lobe that consists of thousands to hundreds of thousands of lobules depending on the species. A second strand (6), (7) addresses image analysis, which in most cases forms the entrance to modeling as it provides the data necessary to generate model hypotheses and to parameterize a model.
(1) Cell types: In a former work by Hoehme et. al. ([113]) a model of liver regeneration after drug-induced damage was established considering hepatocytes and blood vessels. This model has now been expanded to include all relevant cell types, including hepatocytes, blood vessels, hepatic stellate cells, Kupffer cells, invading macrophages and other immune cells. Thereby it is now possible to study perturbations in the temporal scenario of damage and regeneration after signaling events or cells types are knocked down individually or collectively. This model is currently compared to respective perturbation experiments.
(2) Liver disease: Degenerative liver diseases such as liver fibrosis and cirrhosis develop out of a disturbed balance of degenerative and regenerative processes. The model under (1) has thereby been extended by the formation of extracellular matrix, mimicked as fiber networks, to capture the disease process leading to liver fibrosis. In that process characteristic streets form that modify the mechanics, perfusion behavior and detoxification capacity of the liver.

(3) Consequence of liver fibrosis: Whole-slide scans from fibrotic liver in a mouse model has been analysed at different time points after emergence of the disease with regard to the degree of excess matrix to mimic the possible consequences of fibrotic inclusions on perfusion and function of liver within a multiscale model that considers ammonia detoxification in each individual hepatocyte as well as blood flow and transport processes in the liver lobule.
(4) Bile flux: Bile flux has been for decades believed to be controlled by convection at the level of liver lobules as well as at the level of the entire organ. By a methodology based on correlative imaging for quantitative intravital flux analysis no directed advection was detectable in bile canaliculi at the resolution limit. Instead, after active transport across hepatocyte membranes bile salts within the liver lobules are transported in the canaliculi by a diffusion-dominated process. Only in the interlobular ducts i.e., at super-lobular level, diffusion is augmented by advection. In silico simulations of bile transport in real 3D bile network microarchitectures can quantitatively explain the data assuming diffusive transport as sole mechanism.
(5) Liver regeneration after partial hepatectomy (partial organ removal): Partial hepatectomy is an adequate therapy in case of diseases or events that destructed only part of the liver. A typical case is a primary tumor or a metastasis affecting only a single liver lobe. Within an biophysical agent-based model capturing many aspects of the cell mechanics we studied regrowth of liver after partial organ removal in mouse calibrated with multivariate experimental data. Our model predicts characteristic proliferation pattern that change from small animals (as mouse) to large animals (as pig).
(6) Bile duct ligation: Bile duct ligation (BDL) is an experimental procedure that mimics obstructive cholestatic disease. One of the early consequences of BDL in rodents is the appearance of so-called bile infarcts that correspond to Charcot-Gombault necrosis in human cholestasis. The mechanisms causing bile infarcts and their pathophysiological relevance are unclear. Therefore, intravital two photon–based imaging of BDL mice was performed with fluorescent bile salts (BS) and non-BS organic anion analogues. Key findings were followed up by matrix-assisted laser desorption ionization imaging, clinical chemistry, immunostaining, and gene expression analyses. Our group performed analysis of intravital imaging. The key finding is that bile microinfarcts occur in the acute phase after BDL in a limited number of dispersed hepatocytes followed by larger infarcts involving neighboring hepatocytes, and they allow leakage of bile from the BS-overloaded biliary tract into blood, thereby protecting the liver from BS toxicity; in the chronic phase after BDL, reduced sinusoidal BS uptake is a dominant protective factor, and the kidney contributes to the elimination of BS until cholemic nephropathy sets in [108].
(7) Periportalisation during liver fibrosis formation: Within a liver lobule, the function of hepatocytes is zonated i.e., certain functions are only executed by either hepatocytes close to the center (pericentral region) or hepatocytes in the periphery of the lobule (periportal region). Little is known about how liver fibrosis influences lobular zonation. To address this question, three mouse models of liver fibrosis were used, CCl4 administration repeated for 2, 6 and 12 months to induce pericentral damage, as well as bile duct ligation (21 days) and a particular mdr2-mouse model to study periportal fibrosis. Analyses were performed by RNA-sequencing, immunostaining of zonated proteins and image analysis. Image analysis was performed by our group. The key result was that liver fibrosis leads to strong alterations of lobular zonation, where the pericentral region adopts periportal features. Beside adverse consequences, periportalization supports adaptation to repeated doses of hepatotoxic compounds [17].

Toxicity extrapolation from in vitro to in vivo

In vivo toxicity prediction from in vitro data is a major objective in toxicology as it permits bypassing animal experiments, and as the predictive power of animal experiments for human is limited. Objective was the prediction of paracetamol (acetaminophen)-induced hepatotoxicity from in vitro experiments. For this purpose, numerous iterations between in vitro experiments, in vivo experiments and simulations were performed for mouse. Using a recent thesis (Géraldine Cellière'ns PhD thesis [88]) as a start point, two candidate mechanisms could be identified both explaining the in vivo data after calibration of the in silico model with in vitro toxicity data.

Relating imaging on microscopic scales with imaging on macroscopic scales: From Diffusion-Weighted MRI Calibrated With Histological Data: an Example From Lung Cancer

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment, reflecting Brownian movement of water molecules in tissues. Since densely packed cells restrict molecule mobility, tumor tissues produce usually higher signal (less attenuated signal) on isotropic maps compared with normal tissues. However, no general quantitative relation between DWI data and the cell density has been established. In order to link low-resolution clinical cross-sectional data with high resolution histological information, we developed an image processing and analysis chain, which was used to study the correlation between the diffusion coefficient (D value) estimated from DWI and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor. Color deconvolution followed by cell nuclei segmentation was performed on digitized histological images to determine local and cell-type specific 2d (two-dimensional) densities. From these, the 3d cell density was inferred by a model-based sampling technique, which is necessary for the calculation of local and global 3d tumor cell count. Next, DWI sequence information was overlaid with high resolution CT data and the resected histology using prominent anatomical hallmarks for co-registration of histology tissue blocks and non-invasive imaging modalities' data. The integration of cell numbers information and DWI data derived from different tumor areas revealed a clear negative correlation between cell density and D value. Importantly, spatial tumor cell density can be calculated based on DWI data. In summary, our results demonstrate that tumor cell count and heterogeneity can be predicted from DWI data, which may open new opportunities for personalized diagnosis and therapy optimization [157]. The work of that paper has been further advanced to adapt the procedures for clinical use (in preparation).

Collaborations

- Biological control of arboviroses: **Nicolas Vauchelet** (Université Paris 13); **Grégoire Nadin** (LJLL, Sorbonne Université); **Yannick Privat** (Université de Strasbourg); **D. Villela, C. Struchiner** (Fiocruz, Brazil); **Jorge Zubelli** (IMPA, Brazil); **Alain Rapaport** (INRA-Montpellier), **Y. Dumont** (CIRAD-Montpellier); **Ch. Schaerer, P. Pérez-Estigarribia** (UNA, Paraguay), **O. Vasilieva** (Universidad del Valle, Cali, Colombia), **D. Cardona-Salgado** (Universidad Autónoma de Occidente, Cali, Colombia).
- Protein aggregation in amyloid diseases: **Human Rezaei**'s team at Inra Jouy-en-Josas (France) and **W-F Xue**'s team in at university of Kent (Great Britain); **Tom Banks** at the North Carolina State University (USA) and **Philippe Moireau** (M3DISIM)
- Bacterial growth and division: **Lydia Robert**, Sorbonne Université (France)
- Liver research & toxicology: **JG. Hengstler** group (IfADo, Dortmund, Germany); **R. Gebhardt** (Univ. Leipzig); **U. Klingmueller** (DKFZ, Heidelberg); **Irène Vignon-Clementel** (Inria, COMMEDIA)
- Growth in capsules and biomechanics: **Pierre Nassoy**, Institut dOptique Graduate School, Talence, France; **Josef Kaes**, Peter Debye Institute for Soft Matter Physics, Physics, Univ. Leipzig, Germany.
- Wound healing: **Patrizia Bagnerini** (Genova, Numerical methods), **Benoît Ladoux** (Institut Jacques Monod et Mechanobiology Institute Singapore, Biophysics) and **Antonio Jacinto** (CEDOC, Lisbon, Biology and Medicine). (Adipose tissue regeneration) team of **L. Casteilla** (StromaLab, Toulouse). (Axolotl regeneration) team of O. Chara, SysBio group, Argentina.
- Diffusion of morphogen: Center for Computational and Integrative Biology, Rutgers University (Camden, New Jersey), joint work with Professor Nir Yakoby's Drosophila Laboratory
- Linking micro and macro-image information: Oliver Sedlaczek, Univ. and DKFZ Heidelberg, Kai Breuhahn, Univ. Heidelberg.

MONC Project-Team

4. Application Domains

4.1. Tumor growth monitoring and therapeutic evaluation

Each type of cancer is different and requires an adequate model. More specifically, we are currently working on the following diseases:

- Glioma (brain tumors) of various grades,
- Metastases to the lung, liver and brain from various organs,
- Soft-tissue sarcoma,
- Kidney cancer and its metastases,
- non small cell lung carcinoma.

In this context our application domains are

- Image-driven patient-specific simulations of tumor growth and treatments,
- Parameter estimation and data assimilation of medical images.
- Machine and deep learning methods for delineating the lesions and stratifying patients according to their responses to treatment or risks of relapse.

4.2. Biophysical therapies

- Modeling of electrochemotherapy on biological and clinical scales.
- Evaluation of radiotherapy and radiofrequency ablation.

4.3. In-vitro and animals experimentations in oncology

- Theoretical biology of the metastatic process: dynamics of a population of tumors in mutual interactions, dormancy, pre-metastatic and metastatic niche, quantification of metastatic potential and differential effects of anti-angiogenic therapies on primary tumor and metastases.
- Mathematical models for preclinical cancer research: description and prediction of tumor growth and metastatic development, effect of anti-cancerous therapies.

NUMED Project-Team (section vide)

REO Team

4. Application Domains

4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport though the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 ml.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments.

4.3. Cardiac electrophysiology

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [30], [34], [33] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). One of the goal is to use our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macs/M3disym project-team, we are interested in the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology, in particular the simulation of experimental measurement of the field potential of cardiac stem cells in multi-electrode arrays.

SISTM Project-Team

4. Application Domains

4.1. Systems Biology and Translational medicine

Biological and clinical researches have dramatically changed because of the technological advances, leading to the possibility of measuring much more biological quantities than previously. Clinical research studies can include now traditional measurements such as clinical status, but also thousands of cell populations, peptides, gene expressions for a given patient. This has facilitated the transfer of knowledge from basic to clinical science (from "bench side to bedside") and vice versa, a process often called "Translational medicine". However, the analysis of these large amounts of data needs specific methods, especially when one wants to have a global understanding of the information inherent to complex systems through an "integrative analysis". These systems like the immune system are complex because of many interactions within and between many levels (inside cells, between cells, in different tissues, in various species). This has led to a new field called "Systems biology" rapidly adapted to specific topics such as "Systems Immunology" [53], "Systems vaccinology" [50], "Systems medicine" [42]. From the data scientist point of view, two main challenges appear: i) to deal with the massive amount of data ii) to find relevant models capturing observed behaviors.

4.2. HIV immunotherapies

The management of HIV infected patients and the control of the epidemics have been revolutionized by the availability of highly active antiretroviral therapies. Patients treated by these combinations of antiretrovirals have most often undetectable viral loads with an immune reconstitution leading to a survival which is nearly the same to uninfected individuals [46]. Hence, it has been demonstrated that early start of antiretroviral treatments may be good for individual patients as well as for the control of the HIV epidemics (by reducing the transmission from infected people) [41]. However, the implementation of such strategy is difficult especially in developing countries. Some HIV infected individuals do not tolerate antiretroviral regimen or did not reconstitute their immune system. Therefore, vaccine and other immune interventions are required. Many vaccine candidates as well as other immune interventions (IL7, IL15) are currently evaluated. The challenges here are multiple because the effects of these interventions on the immune system are not fully understood, there are no good surrogate markers although the number of measured markers has exponentially increased. Hence, HIV clinical epidemiology has also entered in the era of Big Data because of the very deep evaluation at individual level leading to a huge amount of complex data, repeated over time, even in clinical trials that includes a small number of subjects.

4.3. Translational vaccinology

Vaccines are one of the most efficient tools to prevent and control infectious diseases, and there is a need to increase the number of safe and efficacious vaccines against various pathogens. However, clinical development of vaccines - and of any other investigational product - is a lengthy and costly process. Considering the public health benefits of vaccines, their development needs to be supported and accelerated. During early phase clinical vaccine development (phase I, II trials, translational trials), the number of possible candidate vaccine strategies against a given pathogen that needs to be down-selected in early clinical development is potentially very large. Moreover, during early clinical development there are most often no validated surrogate endpoints to predict the clinical efficacy of a vaccine strategy based on immunogenicity results that could be used as a consensus immunogenicity endpoint and down-selection criterion. This implies considerable uncertainty about the interpretation of immunogenicity results and about the potential value of a vaccine strategy as it transits through early clinical development. Given the complexity of the immune system and the many unknowns in the generation of a protective immune response, early vaccine clinical development nowadays thus takes advantage of high throughput (or “omics”) methods allowing to simultaneously assess a

large number of response markers at different levels ("multi-omics") of the immune system. This has induced a paradigm shift towards early-stage and translational vaccine clinical trials including fewer participants but with thousands of data points collected on every single individual. This is expected to contribute to acceleration of vaccine development thanks to a broader search for immunogenicity signals and a better understanding of the mechanisms induced by each vaccine strategy. However, this remains a difficult research field, both from the immunological as well as from the statistical perspective. Extracting meaningful information from these multi-omics data and transferring it towards an acceleration of vaccine development requires adequate statistical methods, state-of-the art immunological technologies and expertise, and thoughtful interpretation of the results. It thus constitutes research at the interface between disciplines: data science, immunology and vaccinology. Our main current areas of application here are early phase trials of HIV and Ebola vaccine strategies, in which we participate from the initial trial design to the final data analyses.

XPOP Project-Team

4. Application Domains

4.1. Surface Enhanced Raman Spectroscopy

(joint project with HEGP, AP-HP, and Lip(Sys)2, Université Paris-Saclay)

The objective of this work is to evaluate the feasibility of an evolving technique, surface enhanced Raman spectroscopy (SERS) for the analysis of cytotoxic drug concentration. This technique using silver nanoparticles was applied for quantitative analysis of 5-fluorouracil, one of the most widely used molecules in oncology [8].

In view of the high spectral variability observed between the various repetitions of the experiment, and the observed nonlinear interaction between signal concentration and intensity, nonlinear regression methods that take these variabilities into account have been developed.

4.2. Management of severe trauma

(joint project with the Traumabase group, AP-HP)

Major trauma is defined as any injury that endangers the life or the functional integrity of a person. It has been shown that management of major trauma based on standardized and protocol based care improves prognosis of patients especially for the two main causes of death in major trauma i.e., hemorrhage and traumatic brain injury.

However, evidence shows that patient management even in mature trauma systems often exceeds acceptable time frames, and despite existing guidelines deviations from protocol-based care are often observed. These deviations lead to a high variability in care and are associated with bad outcome such as inadequate hemorrhage control or delayed transfusion. Two main factors explain these observations. First, decision-making in trauma care is particularly demanding, because it requires rapid and complex decisions under time pressure in a very dynamic and multi-player environment characterized by high levels of uncertainty and stress. Second, being a complex and multiplayer process, trauma care is affected by fragmentation. Fragmentation is often the result of loss or deformation of information.

This disruptive influence prevents providers to engage with each other and commit to the care process.In order to respond to this challenge, our program has set the ambitious goal to develop a trauma decision support tool, the TraumaMatrix. The program aims to provide an integrative decision support and information management solution to clinicians for the first 24 hours of major trauma management. This program is divided into three steps.

Based on a detailed and high quality trauma database, Step 1 consists in developing the mathematical tools and models to predict trauma specific outcomes and decisions. This step raises considerable scientific and methodological challenges.

Step 2 will use these methods to apply them to develop in close cooperation with trauma care experts the decision support tool and develop a user friendly and ergonomic interface to be used by clinicians.

Step 3 will further develop the tool and interface and test in real-time its impact on clinician decision making and patient outcome.

4.3. Precision medicine and pharmacogenomics

(joint project with Dassault Systèmes)

Pharmacogenomics involves using an individual's genome to determine whether or not a particular therapy, or dose of therapy, will be effective. Indeed, people's reaction to a given drug depends on their physiological state and environmental factors, but also to their individual genetic make-up.

Precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person. While some advances in precision medicine have been made, the practice is not currently in use for most diseases.

Currently, in the traditional population approach, inter-individual variability in the reaction to drugs is modeled using covariates such as weight, age, sex, ethnic origin, etc. Genetic polymorphisms susceptible to modify pharmacokinetic or pharmacodynamic parameters are much harder to include, especially as there are millions of possible polymorphisms (and thus covariates) per patient.

The challenge is to determine which genetic covariates are associated to some PKPD parameters and/or implicated in patient responses to a given drug.

Another problem encountered is the dependence of genes, as indeed, gene expression is a highly regulated process. In cases where the explanatory variables (genomic variants) are correlated, Lasso-type methods for model selection are thwarted.

There is therefore a clear need for new methods and algorithms for the estimation, validation and selection of mixed effects models adapted to the problems of genomic medicine.

A target application of this project concerns the lung cancer.

EGFR (Epidermal Growth Factor Receptor) is a cell surface protein that binds to epidermal growth factor. We know that deregulation of the downstream signaling pathway of EGFR is involved in the development of lung cancers and several gene mutations responsible for this deregulation are known.

Our objective is to identify the variants responsible for the disruption of this pathway using a modelling approach. The data that should be available for developing such model are ERK (Extracellular signal–regulated kinases) phosphorylation time series, obtained from different genetic profiles.

The model that we aim to develop will describe the relationship between the parameters of the pathway and the genomic covariates, i.e. the genetic profile. Variants related to the pathway include: variants that modify the affinity binding of ligands to receptors, variants that modify the total amount of protein, variants that affect the catalytic site,...

4.4. Oncology

(joint project with the Biochemistry lab of Ecole Polytechnique and Institut Curie)

In cancer, the most dreadful event is the formation of metastases that disseminate tumor cells throughout the organism. Cutaneous melanoma is a cancer, where the primary tumor can easily be removed by surgery. However, this cancer is of poor prognosis; because melanomas metastasize often and rapidly. Many melanomas arise from excessive exposure to mutagenic UV from the sun or sunbeds. As a consequence, the mutational burden of melanomas is generally high

RAC1 encodes a small GTPase that induces cell cycle progression and migration of melanoblasts during embryonic development. Patients with the recurrent P29S mutation of RAC1 have 3-fold increased odds at having regional lymph nodes invaded at the time of diagnosis. RAC1 is unlikely to be a good therapeutic target, since a potential inhibitor that would block its catalytic activity, would also lock it into the active GTP-bound state. This project thus investigates the possibility of targeting the signaling pathway downstream of RAC1.

XPOP is mainly involved in Task 1 of the project: *Identifying deregulations and mutations of the ARP2/3 pathway in melanoma patients*.

Association of over-expression or down-regulation of each marker with poor prognosis in terms of invasion of regional lymph nodes, metastases and survival, will be examined using classical univariate and multivariate analysis. We will then develop specific statistical models for survival analysis in order to associate prognosis factors to each composition of complexes. Indeed, one has to implement the further constraint that each subunit has to be contributed by one of several paralogous subunits. An original method previously developed by XPOP has already been successfully applied to WAVE complex data in breast cancer.

The developed models will be rendered user-friendly though a dedicated Rsoftware package.

This project can represent a significant step forward in precision medicine of the cutaneous melanoma.

4.5. Anesthesiology

(joint project with AP-HP Lariboisière and M3DISIM)

Two hundred million general anaesthesias are performed worldwide every year. Low blood pressure during anaesthesia is common and has been identified as a major factor in morbidity and mortality. These events require great reactivity in order to correct them as quickly as possible and impose constraints of reliability and reactivity to monitoring and treatment.

Recently, studies have demonstrated the usefulness of noradrelanine in preventing and treating intraoperative hypotension. The handling of this drug requires great vigilance with regard to the correct dosage. Currently, these drugs are administered manually by the healthcare staff in bolus and/or continuous infusion. This represents a heavy workload and suffers from a great deal of variability in order to find the right dosage for the desired effect on blood pressure.

The objective of this project is to automate the administration of noradrelanine with a closed-loop system that makes it possible to control the treatment in real time to an instantaneous blood pressure measurement.

4.6. Intracellular processes

(joint project with the InBio and IBIS inria teams and the MSC lab, UMR 7057)

Significant cell-to-cell heterogeneity is ubiquitously-observed in isogenic cell populations. Cells respond differently to a same stimulation. For example, accounting for such heterogeneity is essential to quantitatively understand why some bacteria survive antibiotic treatments, some cancer cells escape drug-induced suicide, stem cell do not differentiate, or some cells are not infected by pathogens.

The origins of the variability of biological processes and phenotypes are multifarious. Indeed, the observed heterogeneity of cell responses to a common stimulus can originate from differences in cell phenotypes (age, cell size, ribosome and transcription factor concentrations, etc), from spatio-temporal variations of the cell environments and from the intrinsic randomness of biochemical reactions. From systems and synthetic biology perspectives, understanding the exact contributions of these different sources of heterogeneity on the variability of cell responses is a central question.

The main ambition of this project is to propose a paradigm change in the quantitative modelling of cellular processes by shifting from mean-cell models to single-cell and population models. The main contribution of XPOP focuses on methodological developments for mixed-effects model identification in the context of growing cell populations [9].

- Mixed-effects models usually consider an homogeneous population of independent individuals. This assumption does not hold when the population of cells (i.e. the statistical individuals) consists of several generations of dividing cells. We then need to account for inheritance of single-cell parameters in this population. More precisely, the problem is to attribute the new state and parameter values to newborn cells given (the current estimated values for) the mother.
- The mixed-effects modelling framework corresponds to a strong assumption: differences between cells are static in time (ie, cell-specific parameters have fixed values). However, it is likely that for any given cell, ribosome levels slowly vary across time, since like any other protein, ribosomes are

produced in a stochastic manner. We will therefore extend our modelling framework so as to account for the possible random fluctuations of parameter values in individual cells. Extensions based on stochastic differential equations will be investigated.

- Identifiability is a fundamental prerequisite for model identification and is also closely connected to optimal experimental design. We will derive criteria for theoretical identifiability, in which different parameter values lead to non-identical probability distributions, and for structural identifiability, which concerns the algebraic properties of the structural model, i.e. the ODE system. We will then address the problem of practical identifiability, whereby the model may be theoretically identifiable but the design of the experiment may make parameter estimation difficult and imprecise. An interesting problem is whether accounting for lineage effects can help practical identifiability of the parameters of the individuals in presence of measurement and biological noise.

4.7. Population pharmacometrics

(joint project with Lixoft)

Pharmacometrics involves the analysis and interpretation of data produced in pre-clinical and clinical trials. Population pharmacokinetics studies the variability in drug exposure for clinically safe and effective doses by focusing on identification of patient characteristics which significantly affect or are highly correlated with this variability. Disease progress modeling uses mathematical models to describe, explain, investigate and predict the changes in disease status as a function of time. A disease progress model incorporates functions describing natural disease progression and drug action.

The model based drug development (MBDD) approach establishes quantitative targets for each development step and optimizes the design of each study to meet the target. Optimizing study design requires simulations, which in turn require models. In order to arrive at a meaningful design, mechanisms need to be understood and correctly represented in the mathematical model. Furthermore, the model has to be predictive for future studies. This requirement precludes all purely empirical modeling; instead, models have to be mechanistic.

In particular, physiologically based pharmacokinetic models attempt to mathematically transcribe anatomical, physiological, physical, and chemical descriptions of phenomena involved in the ADME (Absorption - Distribution - Metabolism - Elimination) processes. A system of ordinary differential equations for the quantity of substance in each compartment involves parameters representing blood flow, pulmonary ventilation rate, organ volume, etc.

The ability to describe variability in pharmacometrics model is essential. The nonlinear mixed-effects modeling approach does this by combining the structural model component (the ODE system) with a statistical model, describing the distribution of the parameters between subjects and within subjects, as well as quantifying the unexplained or residual variability within subjects.

The objective of XPOP is to develop new methods for models defined by a very large ODE system, a large number of parameters and a large number of covariates. Contributions of XPOP in this domain are mainly methodological and there is no privileged therapeutic application at this stage [7], [21], [14].

However, it is expected that these new methods will be implemented in software tools, including MONOLIX and Rpackages for practical use.

4.8. Mass spectrometry

(joint project with the Molecular Chemistry Laboratory, LCM, of Ecole Polytechnique)

One of the main recent developments in analytical chemistry is the rapid democratization of high-resolution mass spectrometers. These instruments produce extremely complex mass spectra, which can include several hundred thousand ions when analyzing complex samples. The analysis of complex matrices (biological, agri-food, cosmetic, pharmaceutical, environmental, etc.) is precisely one of the major analytical challenges of this new century. Academic and industrial researchers are particularly interested in trying to quickly and effectively establish the chemical consequences of an event on a complex matrix. This may include, for

example, searching for pesticide degradation products and metabolites in fruits and vegetables, photoproducts of active ingredients in a cosmetic emulsion exposed to UV rays or chlorination products of biocides in hospital effluents. The main difficulty of this type of analysis is based on the high spatial and temporal variability of the samples, which is in addition to the experimental uncertainties inherent in any measurement and requires a large number of samples and analyses to be carried out and computerized data processing (up to 16 million per mass spectrum).

A collaboration between XPOP and the Molecular Chemistry Laboratory (LCM) of the Ecole Polytechnique began in 2018. Our objective is to develop new methods for the statistical analysis of mass spectrometry data.

These methods are implemented in the SPIX software.