



RESEARCH CENTER

FIELD

**Digital Health, Biology and Earth**

# Activity Report 2017

## Section Application Domains

Edition: 2018-02-19



## COMPUTATIONAL BIOLOGY

1. ABS Project-Team (section vide)	5
2. AMIBIO Team	6
3. BEAGLE Project-Team (section vide)	7
4. BIGS Project-Team	8
5. BONSAI Project-Team	9
6. CAPSID Project-Team	10
7. DYLISS Project-Team	12
8. ERABLE Project-Team	14
9. GENSCALE Project-Team	15
10. IBIS Project-Team (section vide)	16
11. LIFEWARE Project-Team	17
12. MORPHEME Project-Team (section vide)	18
13. PLEIADE Team	19
14. SERPICO Project-Team	21
15. TAPDANCE Team (section vide)	23
16. VIRTUAL PLANTS Project-Team (section vide)	24

## COMPUTATIONAL NEUROSCIENCE AND MEDICINE

17. ARAMIS Project-Team	25
18. ASCLEPIOS Project-Team (section vide)	26
19. ATHENA Project-Team	27
20. BIOVISION Team	28
21. CAMIN Team (section vide)	29
22. GALEN Project-Team	30
23. MATHNEURO Team (section vide)	32
24. MIMESIS Team	33
25. MNEMOSYNE Project-Team	34
26. NEUROSYS Project-Team	35
27. PARIETAL Project-Team	37
28. VISAGES Project-Team	39

## EARTH, ENVIRONMENTAL AND ENERGY SCIENCES

29. AIRSEA Project-Team	40
30. ANGE Project-Team	41
31. CASTOR Project-Team	44
32. COFFEE Project-Team	45
33. FLUMINANCE Project-Team	47
34. LEMON Team	48
35. MAGIQUE-3D Project-Team	51
36. SERENA Project-Team	53
37. STEEP Project-Team	54
38. TONUS Team	58

## MODELING AND CONTROL FOR LIFE SCIENCES

39. BIOCORE Project-Team .....	59
40. CARMEN Project-Team .....	61
41. DRACULA Project-Team (section vide) .....	62
42. M3DISIM Project-Team .....	63
43. MAMBA Project-Team (section vide) .....	64
44. MONC Project-Team .....	65
45. MYCENAE Project-Team .....	66
46. NUMED Project-Team (section vide) .....	68
47. REO Project-Team .....	69
48. SISTM Project-Team .....	71
49. XPOP Project-Team .....	72

**ABS Project-Team (section vide)**

## AMIBIO Team

# 4. Application Domains

## 4.1. Circular RNAs

**Participants:** Mireille Régnier, Alice Héliou.

Circular RNAs (circRNAs) have been found abundantly in human cells as well as in many other animals. These non-coding RNAs are involved in the regulation of numerous biological processes, and it was recently shown that, as pre-miRNA, they might actually encode short functional peptides. Our collaborators at Ecole Polytechnique (Biology Dept, LOB) have demonstrated the role of RNA ligase *Pab1020* in RNA circularization. The protein *Pab1020* is a member of the conserved *Rnl3* family of RNA ligases that are predominantly found in hyperthermophiles (archaea, bacteria) and halophiles.

Many computational methods have been proposed to identify and characterize circular RNA from high throughput sequencing data. However, they all suffer from a low specificity, leading to an explosion of false positives. Along with our partners at LOB (Ecole Polytechnique), we develop a robust method for the detection of circRNAs, particularly well-suited to accomodate to analyze sequencing data acquired in extreme environments.

## 4.2. Analysis of probing data

**Participants:** Yann Ponty, Mireille Régnier, Afaf Saaidi.

SHAPE probing [47] is an experimental technique in which RNA is exposed to a reagent which, upon reverse-transcription, induces a modification (truncation, mutation) in the DNA. The prevalence of such modifications, which depends on the locally adopted structure(s) (or lack thereof), can be measured for each nucleotide using sequencing techniques, informing regarding the 2D structure. SHAPE probing data can thus be used by structure prediction methods, either to assess their consistency with a proposed structural model, or to restrict the conformation space.

As part of a colloboration with B. Sargueil's lab (Faculté de pharmacie, Paris V) funded by the Fondation pour la Recherche medical, we strive to propose a new paradigm for the analysis data produced using a new experimental technique, called SHAPE analysis (Selective 2'-Hydroxyl Acylation analyzed by Primer Extension). This experimental setup produces an accessibility profile associated with the different positions of an RNA, the *shadow* of an RNA. We currently design new algorithmic strategies to infer the secondary structure of RNA from multiple SHAPE experiments performed by experimentalists at Paris V. Those are obtained on mutants, and will be coupled with a fragment-based 3D modeling strategy developed by our partners at McGill.

**BEAGLE Project-Team (section vide)**

## **BIGS Project-Team**

# **4. Application Domains**

## **4.1. Tumor growth-oncology**

On this subject, we have new collaborations with clinicians and we want to propose branching processes to model appearance of mutations in tumor. 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 subject of the ITMO project. Another subject is the identification of dynamic network of expression. We continue our work on low-grade gliomas. The ongoing collaboration with Montpellier CHU, and a new one with Montreal CRHUM should provide us more data. We initiate as well interactions with researchers from Montreal LIO to extend the previous work. We still have much work to do in modeling to reach our goal of a decision-aid tool for personalised medicine. In the same context, there is a question of clustering analysis of a brain cartography obtained by sensorial simulations during awake surgery.

## **4.2. Genomic data and micro-organisms population study**

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 through 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 and scoring, of adverse events. We also want to develop our expertise in rupture detection in a project with APHP 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 (during 30mn), 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**

The 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. Through a beginning collaboration with Pr A. Benetos, geriatrician at CHU Nancy, we recently data on the distribution of the length of telomeres from blood cells. With some 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).



## **BONSAI Project-Team**

# **4. Application Domains**

## **4.1. Life Sciences and health**

Our research plays a pivotal role in all fields of life sciences and health where genomic data are involved. This includes more specifically the following topics: plant genomics (genome structure, evolution, microRNAs), cancer (leukemia, mosaic tumors), drug design (NRPSs), environment (metagenomics and metatranscriptomics), virology (evolution, RNA structures) ...

## CAPSID Project-Team

# 4. Application Domains

## 4.1. Biomedical Knowledge Discovery

**Participants:** Marie-Dominique Devignes [contact person], David Ritchie, Sabeur Aridhi, Gabin Personeni, Seyed Ziaeddin Alborzi, Bishnu Sarker, Claire Lacomblez.

This projects in this domain are carried out in collaboration with the Orpailleur Team.

Huge and ever increasing amounts of biomedical data (“Big Data”) are bringing new challenges and novel opportunities for knowledge discovery in biomedicine. We are actively collaborating with biologists and clinicians to design and implement approaches for selecting, integrating, and mining biomedical data in various areas. In particular, we are focusing on leveraging bio-ontologies at all steps of this process (the main thesis topic of Gabin Personeni, co-supervised by Marie-Dominique Devignes and Adrien Coulet from the Orpailleur team). One specific application concerns exploiting Linked Open Data (LOD) to characterise the genes responsible for intellectual deficiency. This work is in collaboration with Pr. P. Jonveaux of the Laboratoire de Génétique Humaine at CHRU Nancy [54], [55]. This involves using inductive logic programming as a machine learning method and at least three different ontologies (Gene Ontology, Human Phenotype Ontology, and Disease Ontology). This approach has also been applied using pattern structure mining (an extension of formal concept analysis) of drug and disease ontologies to discover frequently associated adverse drug events in patients [20]. This work was performed in collaboration with the Centre for BioMedical Informatics Research (BMIR) at Stanford University.

Recently, a new application for biomedical knowledge discovery has emerged from the ANR “FIGHT-HF” (fight heart failure) project, which is in collaboration with several INSERM teams at CHRU Nancy. In this case, the molecular mechanisms that underly HF at the cellular and tissue levels will be considered against a background of all available data and ontologies, and represented in a single integrated complex network. A network platform is under construction with the help of a young start-up company called Edgeleap. Together with this company, we are developing query and analysis facilities to help biologists and clinicians to identify relevant biomarkers for patient phenotyping [34]. Docking of small molecules on candidate receptors, as well as protein-protein docking will also be used to clarify a certain number of relations in the complex HF network.

## 4.2. Prokaryotic Type IV Secretion Systems

**Participants:** Marie-Dominique Devignes [contact person], Bernard Maigret, Isaure Chauvot de Beauchêne, David Ritchie.

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 [26]. 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 [40], [63]. However, the detailed nature of the interactions between the remaining components and those of the core channel remains to be resolved. Therefore, these secretion systems represent another family of complex biological systems (scales 2 and 3) that call for integrated modeling approaches to fully understand their machinery.

In the frame of the Lorraine Université d’Excellence (LUE) “CITRAM” project MD Devignes is pursuing her collaboration with Nathalie Leblond of the Genome Dynamics and Microbial Adaptation (DynAMic) laboratory (UMR 1128, Université de Lorraine, INRA) on the discovery of new 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* genome. As these elements encode all or a subset of the components of the Type IV secretion system, they constitute a valuable source of

sequence data and constraints for modeling these systems in 3D. Another interesting aspect of this particular system is that unlike other secretion systems, the Type IV secretion systems are not restricted to a particular group of bacteria [14].

### 4.3. Protein-RNA Interactions

**Participants:** Isaure Chauvot de Beauchêne [contact person], Bernard Maigret, Maria Elisa Ruiz Echarte, David Ritchie.

As well as playing an essential role in the translation of DNA into proteins, RNA molecules carry out many other essential biological functions in cells, often through their interactions with proteins. A critical challenge in modelling such interactions computationally is that the RNA is often highly flexible, especially in single-stranded (ssRNA) regions of its structure. These flexible regions are often very important because it is through their flexibility that the RNA can adjust its 3D conformation in order to bind to a protein surface. However, conventional protein-protein docking algorithms generally assume that the 3D structures to be docked are rigid, and so are not suitable for modeling protein-RNA interactions. There is therefore much interest in developing protein-RNA docking algorithms which can take RNA flexibility into account.

We are currently developing a novel flexible docking algorithm which first docks small fragments of ssRNA (typically three nucleotides at a time) onto a protein surface, and then combinatorially reassembles those fragments in order to recover a contiguous ssRNA structure on the protein surface [33], [32]. We have since implemented a prototype “forward-backward” dynamic programming algorithm with stochastic backtracking that allows us to model protein RNA interactions for ssRNAs of up to 7 nucleotides without requiring any prior knowledge of the interaction, while still avoiding a brute-force search. In the frame of our PEPS collaboration “InterANRIL” with the IMoPA lab (Univ de Lorraine), we are currently working with biologists to apply the approach to modeling certain long non-coding RNA (lncRNA) complexes. We next plan to build a large library of RNA fragments in order to extend this approach to partially structured RNA molecules, and in the longer term we aim to apply this approach to modeling flexible peptide-protein interactions in a similar way.

## DYLISS Project-Team

## 4. Application Domains

### 4.1. Application fields in biology

Our methods are applied in several fields of molecular biology.

Our main application field is **marine biology**, as it is a transversal field with respect to issues in integrative biology, dynamical systems and sequence analysis. Our main collaborators work at the Station Biologique de Roscoff. We are strongly involved in the study of brown algae: the *meneco*, *memap* and *memerge* tools were designed to realize a complete reconstruction of metabolic networks for non-benchmark species [97], [69]. On the same application model, the pattern discovery tool *protomata learner* combined with supervised bi-clustering based on formal concept analysis allows for the classification of sub-families of specific proteins [67]. The same tool also allowed us to gain a better understanding of cyanobacteria proteins [4]. At the larger level of 4D structures, classification technics have also allowed us to introduce new methods for the characterization of viruses in marine metagenomic sample [72]. Finally, in dynamical systems, we use asymptotic analysis (tool *pogg*) to decipher the initiation of sea urchin translation [51] [95]. We are currently involved in two new applications in this domain: the team participates to a Inria Project Lab program with the Biocore and Ange Inria teams, focused on the understanding on green micro-algae; and we are involved in the deciphering of phytoplankton variability at the system biology level in collaboration with the Station Biologique de Roscoff (ANR Samosa).

In **micro-biology**, our main issue is the understanding of bacteria living in extreme environments, mainly in collaboration with the group of bioinformatics at Universidad de Chile (funded by CMM, CRG and Inria-Chile). In order to elucidate the main characteristics of these bacteria, we develop efficient methods to identify the main groups of regulators for their specific response in their living environment. To that purpose, we use constraints-based modeling and combinatorial optimization. The integrative biology tools *meneco* *bioquali*, *ingranalysis*, *shogen*, *lombarde* were designed in this context [77]. In 2016, two applications focused on the study of extremophile consortium of bacteria have been performed with these tools [61], [38]. In parallel, in collaboration with Ifremer (Brest), we have conducted similar work to decipher protein-protein interactions within archeobacteria [96]. Our sequence analysis tool (*logol*) allowed us to build and maintain a very expressive CRISPR database [10] [50].

Similarly, in **environmental sciences**, our goal is to propose methods to identify regulators of very complex phenotypes related to environmental issues. In collaboration with researchers from Inra/Pegase laboratory, we develop methods to distinguish the response of breeding animals to different diaries or treatments [37] and characterize upstream transcriptional regulators [59], applied to porks [74], [75], [76]. Semantic-based analysis was useful for interpreting differences of gene expression in pork meat [79].

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 [85], [120]. This was performed in collaboration with Inra/Igepp. This paved the way to the recent research track initiated in the team about integration of heterogeneous data with RDF-technologies (see AskOmics software) [84], [70] and about graph-compression (see powergrasp software).

In **bio-medical applications**, we focus our attention on the confrontation of large-scale measurements with large-scale knowledge repositories about regulation pathways such as Transpath, PID or pathway commons. In collaboration with Institut Curie, we have studied the Ewing Sarcoma regulation network to test the capability of our tool *bioquali* to accurately correct and predict a large-scale network behavior [47]. Our ongoing studies in this field focus on the exhaustive learning of discrete dynamical networks matching with experimental data, as a case study for modeling experimental design with constraints-based approaches. To that purpose, we collaborate with J. Saez Rodriguez group at EBI [115] and N. Theret group at Inserm/Irset (Rennes) [40]. The dynamical system tools *caspo* and *cadbiom* were designed within these collaborations. Ongoing studies

focus on the understanding of the metabolism of xenobiotics (mecagenotox program) and the filtering of sets of regulatory compounds within large-scale signaling network (TGFSysBio project).

## **ERABLE Project-Team**

# **4. Application Domains**

## **4.1. Biology & 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. Today the bioinformatics analysis is mainly based on the detection of SNPs (Single Nucleotide Polymorphism) from a set of predefined target genes. Tomorrow, due to the decreasing cost of the sequencing process, bioinformatics analysis will scan the complete genome and report all kinds of mutations, including complex mutations such as large insertions or deletions, that could be associated with cancers.

**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 anatomical 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 and Environment

**Improving plant breeding:** such projects aim at 1) identifying favorable alleles at loci contributing to phenotypic variation, 2) characterizing N-traits at the functional level and 3) 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.

**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.

**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 their role, for example, in the CO<sub>2</sub> 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. However, one important application of our research is the development of a **modeling software** for computational systems biology.

## 4.2. Modeling software for systems 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. Couplings between 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 characteristics. 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<sup>0</sup>. 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 biosensors in non-living vesicles for medical applications. 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.

---

<sup>0</sup>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.

**MORPHEME Project-Team (section vide)**

## PLEIADE 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 develop 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).

## 4.2. 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 Eukaryotes) 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 grapevine, 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 disassemble 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 [21]. 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 developing graph-based models of co-occurrence 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 addressed 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 [24]. 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 [22]. Several approaches were proposed to correct for paralogy, either by working directly on the sequences issued from mapped reads [22] 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 [15]. Workflows to implement statistical methods that correct for diversity biases estimates now need an implementation for biologists.



These studies are essential to unravel the complexity of the endomembrane system and how different machineries evolve together (e.g. see Fig. 1 ). They help to control cell organization and function at different scales through an integrative workflow of methodological and technological developments.

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 transformation. Our methodological goal is also to link dynamics information obtained through diffraction limited light microscopy, eventually at a time regime compatible with live cell imaging. The overview of ultrastructural organization will be achieved by complementary electron microscopical methods. Image visualization and quantitative analysis are of course important and 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 in blood and organs and eventually the formation of secondary tumors and metastases.

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.

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 molecule trafficking.

**TAPDANCE Team (section vide)**

**VIRTUAL PLANTS Project-Team (section vide)**



## **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 neurodegenerative 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.

**ASCLEPIOS Project-Team (section vide)**

## 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.

#### **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 are developing research collaborations with the Neurelec company in Sophia Antipolis (subsidiary of Oticon Medical) and with the leading EEG software company BESA based in Munich. We collaborate with Nice University Hospital on the usage of BCI-based communication for ALS<sup>0</sup> patients.

---

<sup>0</sup>Nice University Hospital hosts a regional reference center for patients suffering from Amyotrophic Lateral Sclerosis

## BIOVISION Team

# 4. Application Domains

## 4.1. Applications of cross reality technologies for low-vision

- High tech vision aid systems for low-vision patients: Vision aid systems for low-vision patients is an application domain with commercial products already existing. A variety of XR solutions are on the market using different kinds of platforms (dedicated or large public ones) and offering different kinds of functionalities (magnification, enhancement, text to speech, face and object recognition). Our goal is to propose new solutions to help patients in a given task.
- Serious games <sup>0</sup> for rehabilitation: Since cross reality technology is now becoming available for the large public, it is a promising platform to develop new rehabilitation exercises.
- Cognitive research: A new trend in cross reality technology is to include eye-tracking so that these platforms could also be very efficient to conduct cognitive and behavioural research on a large scale.

## 4.2. Applications of vision modeling studies

- Neuroscience research: Making in-silico experiment 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, 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.
- Education: The simulation platform we develop could also be a useful tool for educational purposes, illustrating for students how the retina works and respond to visual stimuli.

---

<sup>0</sup>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.

**CAMIN Team (section vide)**

## GALEN Project-Team

# 4. Application Domains

## 4.1. 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 will be modeled.

## 4.2. Inference of gene regulatory networks

**Participants:** 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 different 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.

## 4.3. Lung Tumor Detection and Characterization

**Participants:** Evgenios Kornaropoulos, Evangelia Zacharaki, Nikos Paragios

The use of Diffusion Weighted MR Imaging (DWI) is investigated as an alternative tool to radiologists for tumor detection, tumor characterization, distinguishing tumor tissue from non-tumor tissue, and monitoring and predicting treatment response. In collaboration with Hôpitaux Universitaires Henri-Mondor in Paris, France and Chang Gung Memorial Hospital – Linkou in Taipei, Taiwan we investigate the use of modelbased methods of 3D image registration, clustering and segmentation towards the development of a framework for automatic interpretation of images, and in particular extraction of meaningful biomarkers in aggressive lymphomas.

## 4.4. Protein function prediction

**Participants:** Evangelia Zacharaki, Nikos Paragios (in collaboration with D. Vlachakis, University of Patras, Greece)

The massive expansion of the worldwide Protein Data Bank (PDB) provides new opportunities for computational approaches which can learn from available data and extrapolate the knowledge into new coming instances. The aim of our work was to exploit experimentally acquired structural information of enzymes through machine learning techniques in order to produce models that predict enzymatic function.

## 4.5. Imaging biomarkers for chronic lung diseases

**Participants:** Guillaume Chassagnon, Evangelia Zacharaki, Maria Vakalopoulou, Nikos Paragios

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 histogram or texture analysis and assessment of the regional lung elasticity through deformable registration. This work is in collaboration with the Department of Radiology, Cochin Hospital, Paris.

## 4.6. Co-segmentation and Co-registration of Subcortical Brain Structures

**Participants:** Enzo Ferrante, Nikos Paragios, Iasonas Kokkinos

New algorithms to perform co-segmentation and co-registration of subcortical brain structures on MRI images were investigated in collaboration with Ecole Polytechnique de Montreal and the Sainte-Justine Hospital Research Center from Montreal. Brain subcortical structures are involved in different neurodegenerative and neuropsychiatric disorders, including schizophrenia, Alzheimers disease, attention deficit, and subtypes of epilepsy. Segmenting these parts of the brain enables a physician to extract indicators, facilitating their quantitative analysis and characterization. We are investigating how estimated maps of semantic labels (obtained using machine learning techniques) can be used as a surrogate for unlabelled data. We are exploring how to combine them with multi-population deformable registration to improve both alignment and segmentation of these challenging brain structures.

## 4.7. Restoration of old video archives

**Participants:** Emilie Chouzenoux, Jean-Christophe Pesquet (collaboration F. Abboud, WITBE, J.-H. Chenot and L. Laborelli, INA)

The last century has witnessed an explosion in the amount of video data stored with holders such as the National Audiovisual Institute whose mission is to preserve and promote the content of French broadcast programs. the cultural impact of these records, their value is increased due to commercial reexploitation through recent visual media. However, the perceived quality of the old data fails to satisfy the current public demand. The purpose of our work is to propose new methods for restoring video sequences supplied from television archive documents, using modern optimization techniques with proven convergence properties.

**MATHNEURO Team (section vide)**



## **MIMESIS Team**

## **4. Application Domains**

### **4.1. Surgical Training**

Virtual training prevents medical students from early manipulation of real patients. The development of simulation used for medical training usually requires important computational power, since realistic behaviours 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 a 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 a 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 Navigation**

Besides the surgery training and planning, another major need from clinicians is surgical guidance. While the practitioner 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 area such as vessels or to highlight the position of a tumour 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.

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 undoubtedly 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. General remarks

The research directions of the team are motivated by general anesthesia that has attracted our attention in the last years. The following paragraphs explain in some detail the motivation of our work on the four major phenomena of general anesthesia: loss of consciousness, immobility, amnesia and analgesia.

During general anesthesia, the electroencephalogram (EEG) on the scalp changes characteristically: increasing the anesthetic drug concentration the amplitudes of oscillations in the  $\alpha$ -band ( $\sim 8 - 12\text{Hz}$ ) and in the  $\delta$ -band ( $2 - 8\text{Hz}$ ) increase amplitudes in frontal electrodes at low drug concentrations whereas the spectral power decreases in the  $\gamma$ -band ( $\sim 20 - 60\text{Hz}$ ). This characteristic change in the power is the basis of today's EEG-monitors that assist the anesthetist in the control of the anesthesia depths of patients during surgery. However, the conventional monitors exhibit a large variability between the detected anesthetic depth and the real depth of patients. Moreover, a certain number of patients re-gain consciousness during surgery (about 1 – 2 out of 1000) and a large percentage of patients suffer from diverse after-effects, such as nausea or long-lasting cognitive impairments such as partial amnesia (from days to weeks). Since surgery under general anesthesia is part of a hospital's everyday practice, a large number of patients suffer from these events every day. One reason for the lacking control of such disadvantageous effects is the dramatic lack of knowledge on what is going on in the brain during general anesthesia and a weak EEG-online monitoring system during anesthesia. Consequently, to improve the situation of patients during and after surgery and to develop improved anesthetic procedures or even drugs, research is necessary to learn more about the neural processes in the brain and develop new monitoring machines.

## 4.2. Level of consciousness

The EEG originates from coherent neural activity of populations in the cortex. Hence to understand better the characteristic power changes in EEG during anesthesia, it is necessary to study neural population dynamics subject to the concentration of anesthetic drugs and their action on receptors on the single neuron level. We study mathematical models which will be constrained by the signal features extracted from experimental data, such as EEG (data provided by Jamie Sleight, university of Auckland and Christoph Destrieux, university of Tours), Local Field Potentials (data provided by Flavio Fröhlich, university of North Carolina - Chapel Hill) and behavior. The combination of model and analysis of experimental data provides the optimal framework to reveal new knowledge on the neural origin of behavioral features, such as the loss of consciousness or the uncontrolled gain of consciousness during surgery. For instance, modeling studies show that the characteristic changes of spectral power (second-order statistics) are not sufficient to deduce all underlying neural mechanisms. Consequently, additional higher-order statistical measures may provide additional insight into underlying neural mechanisms and may provide a novel marker for the loss of consciousness.

Moreover, the constant supervision of anesthetized patients in intensive care is a demanding task for the personnel in hospital practice. It is almost not possible to take care of a patient constantly and hence the today's medicine demands monitoring devices that control automatically the level of anesthetic drugs based on the patients' neural activity (e.g., EEG). Brain-Computer-Interfaces (BCI) have already demonstrated their potential for the detection of consciousness in non-responsive patients. We will apply the data analysis techniques known in BCI to extract new markers for the depth of anesthesia. More specifically, for deeper anesthesia, auditory-evoked and Event-Related Desynchronization/Event-Related Synchronization (ERD/ERS) BCI could be used to better identify the state of consciousness in patients under anesthesia. In this context, we have established a first contact to the university of Würzburg. Another research direction will link intracranial EEG and scalp EEG by characterising micro-awake episodes during sleep.

### 4.3. Immobility

A research direction will be to take benefit of the relationship between the motor activity and anesthesia. Indeed, even if no movement is visually perceptible, a study by electroencephalographic recordings of brain activity in motor areas, quantifying the characteristics of amplitude and phase synchronization observed in the alpha and beta frequency bands, may reveal an intention of movement. This feature is important because it demonstrates that the patient is aware. Thus, we will develop an experimental protocol in collaboration with an anesthesiologist of the regional hospital on stimulating the median nerve at forearm level to track the evolution of the shape of the beta rebound in the motor cortex for various doses of the anesthetic agent.

### 4.4. Amnesia

Patients sometimes develop post-traumatic disorders associated with the surgery they underwent because they either woke up during the surgery or because the amnesiant effect of the general anesthesia was only partial, declarative memory being maintained in some unexplained cases. It is still unknown how memory can be maintained under general anesthesia and it needs to be investigated to improve the recovery from anesthesia and to avoid as much as possible post-traumatic disorders. To learn more about memory under anesthesia, we will focus our theoretical studies on the oscillation regimes observed in the hippocampus, mainly in the theta and gamma ranges, which are correlated with memory formation and retrieval.

### 4.5. Analgesia

One of the most important aspect in general anesthesia is the loss of pain. During surgery, it is very difficult to find out whether the anesthetized patient feels pain and hence will develop cognitive impairment after surgery. Today, the anesthesiologist knows and detects physiological signs of pain, such as sweat, colour of skin or spontaneous involuntary movements. However, more objective criteria based on EEG may assist the pain detection and hence improve the patients' situation. To this end, we analyze large sets of patient EEG-data observed during surgery and aim to extract EEG signal features of pain.

## 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 electrophysiological 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, Magnetoencephalography (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<sup>0</sup> 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.

---

<sup>0</sup>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<sup>0</sup> 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 [54], and particularly so due to the typically number of subjects included in brain imaging studies (20 to 30, this number tends to increase [55]): 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.

---

<sup>0</sup>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 paid by a factor of 8 in terms of SNR.

## VISAGES Project-Team

# 4. Application Domains

## 4.1. Neuroimaging

One research objective in neuroimaging is the construction of anatomical and functional cerebral maps under normal and pathological conditions. Many researches are currently performed to find correlations between anatomical structures, essentially sulci and gyri, where neuronal activation takes place, and cerebral functions, as assessed by recordings obtained by the means of various neuroimaging modalities, such as PET (Positron Emission Tomography), fMRI (Functional Magnetic Resonance Imaging), EEG (Electro-EncephaloGraphy) and MEG (Magneto-EncephaloGraphy). Then, a central problem inherent to the formation of such maps is to put together recordings obtained from different modalities and from different subjects. This mapping can be greatly facilitated by the use of MR anatomical brain scans with high spatial resolution that allows a proper visualization of fine anatomical structures (sulci and gyri). Recent improvements in image processing techniques, such as segmentation, registration, delineation of the cortical ribbon, modeling of anatomical structures and multi-modality fusion, make possible this ambitious goal in neuroimaging. This problem is very rich in terms of applications since both clinical and neuroscience applications share similar problems. Since this domain is very generic by nature, our major contributions are directed towards clinical needs even though our work can address some specific aspects related to the neuroscience domain.

## 4.2. Multiple sclerosis

Over the past years, a discrepancy became apparent between clinical Multiple sclerosis (MS) classification describing on the one hand MS according to four different disease courses and, on the other hand, the description of two different disease stages (an early inflammatory and a subsequently neurodegenerative phase). It is to be expected that neuroimaging will play a critical role to define in vivo those four different MS lesion patterns. An in vivo distinction between the four MS lesion patterns, and also between early and late stages of MS will have an important impact in the future for a better understanding of the natural history of MS and even more for the appropriate selection and monitoring of drug treatment in MS patients. MRI has a low specificity for defining in more detail the pathological changes which could discriminate between the different lesion types. However, it has a high sensitivity to detect focal and also widespread, diffuse pathology of the normal appearing white and gray matter. Our major objective within this application domain is then to define new neuroimaging markers for tracking the evolution of the pathology from high dimensional data (e.g., nD+t MRI) in the brain and the spinal cord. In addition, in order to complement MR neuroimaging data, we ambition to perform also cell labeling neuroimaging (e.g., MRI or PET) and to compare MR and PET data using standard and experimental MR contrast agents and radiolabeled PET tracers for activated microglia (e.g., USPIO or PK 11195). The goal is to define and develop, for routine purposes, cell specific and also quantitative imaging markers for the improved in vivo characterization of MS pathology.

## 4.3. Modeling of anatomical and anatomo-functional neurological patterns

The major objective within this application domain is to build anatomical and functional brain atlases in the context of functional mapping and for the study of developmental, neurodegenerative or even psychiatric brain diseases (Multiple sclerosis, Epilepsy, Parkinson, Dysphasia, Depression or even Alzheimer). This is a very competitive research domain; our contribution is based on our previous works in this field, and by continuing our local and wider collaborations. An additional objective within this application domain is to find new descriptors to study the brain anatomy and/or function (e.g., variation of brain perfusion, evolution in shape and size of an anatomical structure in relation with pathology or functional patterns, computation of asymmetries ...). This is also a very critical research domain, especially for many developmental or neurodegenerative brain diseases.

## 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 they 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 origins. 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 CO<sub>2</sub> 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.

Practical 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**

# **4. Application Domains**

## **4.1. Magnetic confinement fusion**

The main application domain is magnetic confined fusion. A part of the work is actually used on european Tokamaks like JET, WEST and ITER.

## COFFEE Project-Team

# 4. Application Domains

## 4.1. Porous Media

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.

## 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. Biological degradation, biofilms formation and algae proliferation**

Members of the team have started an original research program devoted to biofilms formation and algae proliferation. We started working on this subject through a collaboration with Roberto Natalini and a group of experts in Firenze interested in preventing damages on historical monuments. It is also motivated by *Ostreopsis* proliferation in the Mediterranean Sea. The multidisciplinary character of this research relies on discussions with researchers of the Oceanography Laboratory in Villefranche-sur-Mer, a leading marine research unit, and the Inria team BIOCORE, led by J-L Gouzé. This research was supported by a ANR-project, led by M. Ribot, and it was the main topic of the PhD thesis of B. Polizzi.

## FLUMINANCE Project-Team

# 4. Application Domains

## 4.1. Introduction

By designing new approaches for the analysis of fluid-image sequences the FLUMINANCE group aims at contributing to several application domains of great interest for the community and in which the analysis of complex fluid flows plays a central role. The group focuses mainly on two broad application domains:

- Environmental sciences;
- Experimental fluid mechanics and industrial flows.

We detail hereafter these two application domains.

## 4.2. Environmental sciences

The first huge application domain concerns all the sciences that aim at observing the biosphere evolution such as meteorology, climatology or oceanography but also remote sensing study for the monitoring of meteorological events or human activities consequences. For all these domains image analysis is a practical and unique tool to *observe, detect, measure, characterize or analyze* the evolution of physical parameters over a large domain. The design of generic image processing techniques for all these domains might offer practical software tools to measure precisely the evolution of fluid flows for weather forecasting or climatology studies. It might also offer possibilities of close surveillance of human and natural activities in sensible areas such as forests, river edges, and valley in order to monitor pollution, floods or fire. The need in terms of local weather forecasting, risk prevention, or local climate change is becoming crucial for our tomorrow's life. At a more local scale, image sensors may also be of major utility to analyze precisely the effect of air curtains for safe packaging in agro-industrial.

## 4.3. Experimental fluid mechanics and industrial flows

In the domain of **experimental fluid mechanics**, the visualization of fluid flows plays a major role, especially for turbulence study since high frequency imaging has been made currently available. Together with analysis of turbulence at different scales, one of the major goals pursued at the moment by many scientists and engineers consists in studying the ability to manipulate a flow to induce a desired change. This is of huge technological importance to enhance or inhibit mixing in shear flows, improve energetic efficiency or control the physical effects of strain and stresses. This is for instance of particular interest for:

- military applications, for example to limit the infra-red signatures of fighter aircraft;
- aeronautics and transportation, to limit fuel consumption by controlling drag and lift effects of turbulence and boundary layer behavior;
- industrial applications, for example to monitor flowing, melting, mixing or swelling of processed materials, or preserve manufactured products from contamination by airborne pollutants, or in industrial chemistry to increase chemical reactions by acting on turbulence phenomena.

## LEMON Team

## 4. Application Domains

### 4.1. Inland flow processes

#### 4.1.1. *Shallow water models with porosity*

Simulating urban floods and free surface flows in wetlands requires considerable computational power. Two-dimensional shallow water models are needed. Capturing the relevant hydraulic detail often requires computational cell sizes smaller than one meter. For instance, meshing a complete urban area with a sufficient accuracy would require  $10^6$  to  $10^8$  cells, and simulating one second often requires several CPU seconds. This makes the use of such model for crisis management impossible. Similar issues arise when modelling wetlands and coastal lagoons, where large areas are often connected by an overwhelming number of narrow channels, obstructed by vegetation and a strongly variable bathymetry. Describing such channels with the level of detail required in a 2D model is impracticable. A new generation of models overcoming this issue has emerged over the last 20 years: porosity-based shallow water models. They are obtained by averaging the two-dimensional shallow water equations over large areas containing both water and a solid phase [35]. The size of a computational cell can be increased by a factor 10 to 50 compared to a 2D shallow water model, with CPU times reduced by 2 to 3 orders of magnitude [53]. While the research on porosity-based shallow water models has accelerated over the past decade [48], [64], [68], [46], [45], [53], [78], [79], [74], [47], a number of research issues remain pending.

#### 4.1.2. *Forcing*

Reproducing optimally realistic spatio-temporal rainfall fields is of salient importance to the forcing of hydrodynamic models. This challenging task requires combining intense, usual and dry weather events. Far from being straightforward, this combination of extreme and non-extreme scenarios requires a realistic modelling of the transitions between normal and extreme periods. [57] have proposed in a univariate framework a statistical model that can serve as a generator and that takes into account low, moderate and intense precipitation. In the same vein, [75] developed a bivariate model. However, its extension to a spatial framework remains a challenge. Existing spatial precipitation stochastic generators are generally based on Gaussian spatial processes [22], [55], that are not adapted to generate extreme rainfall events. Recent advances in spatio-temporal extremes modelling based on generalized Pareto processes [39], [71] and semi-parametric simulation techniques [28] are very promising and could form the base for relevant developments in our framework.

#### 4.1.3. *Inland hydrobiological systems*

Water bodies such as lakes or coastal lagoons (possibly connected to the sea) located in high human activity areas are subject to various kinds of stress such as industrial pollution, high water demand or bacterial blooms caused by freshwater over-enrichment. For obvious environmental reasons, these water resources have to be protected, hence the need to better understand and possibly control such fragile ecosystems to eventually develop decision-making tools. From a modelling point of view, they share a common feature in that they all involve interacting biological and hydrological processes. According to [40], models may be classified into two main types: “minimal dynamic models” and “complex dynamic models”. These two model types do not have the same objectives. While the former are more heuristic and rather depict the likelihood of considered processes, the latter are usually derived from fundamental laws of biochemistry or fluid dynamics. Of course, the latter necessitate much more computational resources than the former. In addition, controlling such complex systems (usually governed by PDEs) is by far more difficult than controlling the simpler ODE-driven command systems.

LEMON has already contributed both to the reduction of PDE models for the simulation of water confinement in coastal lagoons [41], [23] and to the improvement of ODE models in order to account for space-heterogeneity of bioremediation processes in water resources [21].



#### 4.1.4. Parametrization

Numerical modelling requires data acquisition, both for model validation and for parameter assessment. Model benchmarking against laboratory experiments is an essential step and is essential to team's strategy. However, scale model experiments may have several drawbacks: *i*) experiments are very expensive and extremely time-consuming, *ii*) experiments cannot always be replicated, and measurement have precision and reliability limitations, *iii*) dimensional similarity (in terms of geometry and flow characteristic variables such as Froude or Reynolds numbers) cannot always be preserved.

An ideal way to obtain data would be to carry out in situ measurements. But this would be too costly at the scale of studied systems, not to mention the fact that field may become impracticable during flood periods.

Remote sensing data are becoming widely available with high spatial and temporal resolutions. Several recent studies have shown that flood extends can be extracted from optical or radar images [42], for example: to characterize the flood dynamics of great rivers [58], to monitor temporary ponds [69], but also to calibrate hydrodynamics models and assess roughness parameters [66], [49], [77].

Upscaled models developed in LEMON embed new parameters that reflect the statistical properties of the medium geometry. Two types of information are needed: the directional properties of the medium and its flow connectivity properties. New methods are thus to be developed to characterize such statistical properties from geographical data.

## 4.2. Marine and coastal systems

### 4.2.1. Multi-scale ocean modelling

In physical oceanography, all operational models - regardless of the scale they apply to - are derived from the complete equations of geophysical fluid dynamics. Depending on the considered process properties (nonlinearity, scale) and the available computational power, the original equations are adapted with some simplifying hypotheses. The reader can refer to [63], [56] for a hierarchical presentation of such models.

In the nearshore area, the hydrostatic approximation that is used in most large scales models (high sea) cannot be used without a massive loss of accuracy. In particular, shallow water models are inappropriate to describe the physical processes that occur in this zone (see Figure 1 above). This is why Boussinesq-type models are preferred: see [54]. They embed dispersive terms that allow for shoaling and other bathymetry effects. Since the pioneering works of Green and Naghdi (see [43]), numerous theoretical and numerical studies have been delivered by the "mathematical oceanography" community, more specifically in France (see the works of Lannes, Marche, Sainte-Marie, Bresch, etc.). The corresponding numerical models (BOSZ, WaveBox) must thus be integrated in any reasonable nearshore modelling platform.

However, these models cannot simply replace all previous models everywhere in the ocean: dispersive models are useless away from the shore and it is known that wave breaking cannot be simulated using Boussinesq-type equations. Hence the need to couple these models with others. Some work has been done in this direction with a multi-level nesting using software packages such as ROMS, but to the best of our knowledge, all the "boxes" rely on the same governing equations with different grid resolutions. A real coupling between different models is a more difficult task since different models may have different mathematical properties, as shown in the work by Blayo and Rousseau on shallow water modelling (see [1] and [24]).

### 4.2.2. Data-model interactions

An alternative to direct observations is the chaining of numerical models, which for instance represent the physic from offshore to coastal areas. Typically, output data from atmospheric and ocean circulation models are used as forcings for a wave model, which in turn feeds a littoral model. In the case of extreme events, their numerical simulation from physical models is generally unreachable. This is due to a lack of knowledge on boundary conditions and on their physical reliability for such extreme quantities. Based on numerical simulated data, an alternative is to use statistical approaches. [28] proposed such an approach. They first produced and studied a 52-year hindcast using the WW3 wave model [26], [29], [27], [72]. Then stemming from parts of the original work of [25], [44], [39], [28] proposed a semi-parametric approach which aims

to simulate extreme space-time waves processes to, in turn, force a littoral hazard model. Nevertheless their approach allows only a very small number of scenarios to be simulated.

## **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
- geological sequestration of CO<sub>2</sub>
- production of oil and gas

#### **4.2. Complex Stokes and Navier–Stokes flows**

- industrial risks in energy production (fission)

#### **4.3. Energy production, sustainable use of resources**

- simulation of shock waves impinging on deformable or fragmentable structures
- use of nets of rods for sustainable construction

#### **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

## STEPP 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. Let us 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.

STEPP 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<sup>0</sup>, on user behavior<sup>0</sup>, or on the environment<sup>0</sup>. 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

<sup>0</sup> Congestion, cost and time spent for the transport, etc.

<sup>0</sup> Changes in modality choice.

<sup>0</sup> 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<sup>0</sup>. These two models are very different but particularly representative of the main current philosophies and trends in this scientific domain. Their comparison is informative.

**STEPP 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<sup>0</sup>, 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.

---

<sup>0</sup><http://www.urbansim.org>

<sup>0</sup><http://www.aurg.org/>

## TONUS 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 made of 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. Among these 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 for 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 fusion reactions, 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. The applications are electromagnetic 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 new 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 Environmental 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<sub>2</sub> fixation and fluxes**

Phytoplanktonic species, which assimilate CO<sub>2</sub> during photosynthesis, have received a lot of attention in the last years. Microalgal based processes have been developed in order to mitigate industrial CO<sub>2</sub>. 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<sub>2</sub> 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<sub>2</sub> 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 (compared to chemicals)... The use of biocontrol agents, which are, generically, natural enemies (predators, parasitoids or pathogens) of crop pests [66], 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 controlling 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 [71] 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 [53], [52], [45]. 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 Liryc 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 helps 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 [28], 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 arrhythmias 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 [51]. These heterogeneities are 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 and (2) modeling the heterogeneities related to the complex organization and disorganization of the myocytes and fibroblasts. Point (1) is supposed to play a major role in sudden cardiac death and point (2) 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 [25].

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.

**DRACULA Project-Team (section vide)**

## **M3DISIM Project-Team**

# **4. Application Domains**

## **4.1. Clinical applications**

After several validation steps – based on clinical and experimental data – we have reached the point of having validated the heart model in a pre-clinical context where we have combined direct and inverse modeling in order to bring predictive answers on specific patient states. For example, we have demonstrated the predictive ability of our model to set up pacemaker devices for a specific patient in cardiac resynchronization therapies, see [11]. We have also used our parametric estimation procedure to provide a quantitative characterization of an infarct in a clinical experiment performed with pigs, see [1].

**MAMBA Project-Team (section vide)**



## **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,
- EGFR-mutated lung cancer.

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.

## **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.

## MYCENAE Project-Team

## 4. Application Domains

### 4.1. Introduction

MYCENAE addresses rather “upstream” questions in neuroendocrinology and neuroscience. Nevertheless, MYCENAE’s expected results can contribute to more applied issues in these fields, mainly by helping understand the mechanisms underlying physiological and pathological processes and also by designing new concepts for biomedical data analysis. MYCENAE thematics are related to societal issues concerning endocrine disruptors, reproductive biotechnologies, and neurological diseases, especially in case of pathological synchronizations encountered in epilepsy and Parkinson’s disease.

### 4.2. Neuroendocrinology and Neuroscience

We are interested in the complex dynamical processes arising within neuroendocrine axes, with a special focus on the reproductive (hypothalamo-pituitary-gonadal) axis. This axis can be considered as the paragon of neuroendocrine axes, since it both concentrates all remarkable dynamics that can be exhibited by these axes and owns its unique specificities, as gonads are the only organs that host germ cells. Since, in neuroendocrine axes, neural systems are embedded within endocrine feedback loops and interact with peripheral organs, one also needs to get interested in the peripheral dynamics to be able to “close the loop” and account for the effect of peripheral inputs on neural dynamics. In the case of the HPG axis, these dynamics are especially complex, because they involve developmental processes that occur even in adult organisms and combine the glandular function of the gonads with their gametogenic function.

Neuroendocrinology is thus a scientific field at the interface between Neuroscience, Endocrinology and Physiology (and even of Developmental Biology in the case of the HPG axis). On a neuroscience ground, mathematical neuroendocrinology is specifically interested in endocrine neurons, which have the uncommon ability of secreting neurohormones into the blood stream. Neuroendocrine networks are characterized by the emergence of very slow rhythms (on the order of an hour), finite size effects due to their relative small number of neurons (on the order of a few thousands for the Gonadotropin-Releasing-Hormone network) and neuroanatomical particularities, that impact the way they can synchronize and desynchronize. On a physiological ground, gonadal cell biology raises specific cell biology issues on more than one account. First, the gonads are the only organs sheltering the germ cell lines (corresponding to oogenesis in ovaries and spermatogenesis in testes). Hence, the two modes of cell division, mitosis and meiosis are encountered in these tissues. Second, there are intricate interactions between the gonadal somatic cells (granulosa cells in the ovaries, sertoli cells in the testes) and the germ cells. Third, the control of gonadal cell populations is exerted within endocrine feedback loops involving both the hypothalamus and pituitary, which results naturally in multiscale population dynamics coupled with hormonally-controlled cell kinetics.

MYCENAE’s research topics in mathematical neuroscience deal with complex oscillations, synchronization and plasticity.

We study (i) the emergence of network-level behaviors from individual dynamics of excitable cells (mainly neurons, but not exclusively, as the pituitary cells belong to the family of excitable cells): complete synchronization or synchronization of specific events, effect of the recruitment rate in the synchronization process, dependence on the neuro-anatomical and functional coupling properties; (ii) the control of the different possible configurations of the network depending on external (e.g. daylength) and/or internal inputs (e.g. metabolic status), at the source of plasticity processes in cognitive (vision learning) or neuroendocrine systems (differential sensitivity to gonadal steroids and peptides across the different steps of the reproductive life); (iii) the encoding of neuro-hormonal signals as complex oscillations, on the electrical, ionic (calcium dynamics) and secretory levels; and (iv) the decoding of these signals by their target neuronal or non-neuronal cells.

More recently, we have been interested into developmental biology issues in neurosciences: neurogenesis and brain development. The anatomical and functional organization of the nervous system, and especially the brain, is highly structured and tightly regulated. The surface of the cortex, its thickness, but also the size and shape of the brain areas associated to the different sensory or motor areas are very reliable quantities across different individuals. In collaboration with different teams of biologists, we develop and investigate models of the development of the brain, at different time and spatial scale.

The biological relevance of our modeling and model-based signal analysis approaches is grounded on our network of collaborations with teams of experimentalist biologists. In particular, we have long standing collaborations with the UMR 6175 (INRA-CNRS-Université François Rabelais-Haras Nationaux) “Physiologie de la Reproduction et des Comportements” that covers most our research topics in reproductive neuroendocrinology. We have especially close links with the Bingo (Integrative Biology of the ovary) and Bios (Biology and Bioinformatics of Signaling Systems) teams, which were partners of the REGATE LSIA. We have been jointly investigating issues relative to terminal or basal follicular development [7], [8], analysis of neurosecretory patterns [16] and modeling of GPCR (G-Protein Coupled Receptors) signaling networks [10]. We also have special links with the Center for Interdisciplinary Research in Biology (CIRB, Collège de France), headed by Alain Prochiantz, that help us get a better understanding of how the brain connectivity develops and how it is functionally organized. An instance of a recent collaborative work is the study of the organization of spatial frequencies in the primary visual cortex [34].

**NUMED Project-Team (section vide)**

## REO Project-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 through 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.* [50], [55], [54] 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" [55], "Systems vaccinology" [52], "Systems medicine" [44]. From the statistician 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. The case of HIV immunology**

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 [48]. 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) [43]. 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. The case of Ebola vaccine development**

In response to the recent outbreak of Ebola virus disease in West Africa, the clinical development of some candidate to Ebola vaccine has been accelerated. Several vectors, mostly encoding glycoprotein of the virus, were tested in Phase I-II studies in order to assess their safety and immunogenicity. One of the main question of interest there is the antibody response induced by vaccination, as some non-human primates studies have shown protection against the virus when antibody levels were high enough. Although bridging studies still have to be developed, antibodies are thus considered as a criterium of interest. The challenge is then to evaluate the durability of the antibody response, whether it be at an individual or population level, in order to evaluate the impact of a vaccine strategy in case of an epidemic. Moreover, we are interested in the factors associated to this antibody response, and even more the other immune markers (from both innate and adaptive immune response) able to predict antibody levels. As those relationship are non-linear, sophisticated statistical and mathematical methods are developed in order to address these questions. A systems medicine approach using multidimensional immunogenicity data from clinical trials and statistical models can help to understand vaccine mechanisms and improve the selection of optimised vaccine strategies for clinical trials.

## **XPOP Project-Team**

# **4. Application Domains**

## **4.1. Precision medicine and pharmacogenomics**

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,...

## **4.2. Oncology**

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 through a dedicated Rsoftware package.

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

### 4.3. Hemodialysis

Hemodialysis is a process for removing waste and excess water from the blood and is used primarily as an artificial replacement for lost kidney function in people with kidney failure. Side effects caused by removing too much fluid and/or removing fluid too rapidly include low blood pressure, fatigue, chest pains, leg-cramps, nausea and headaches.

Nephrologists must therefore correctly assess the hydration status in chronic hemodialysis patients and consider fluid overload effects when prescribing dialysis, according to a new study.

The fluid overload biomarker, B-type natriuretic peptide (BNP) is an important component of managing patients with kidney disease. Indeed, it is believed that each dialysis patient will have an ideal or "dry" BNP level which will accurately and reproducibly reflect their optimal fluid status.

The objective of this study is to develop a model for the BNP and the hydration status using individual information (age, sex, ethnicity, systolic blood pressure, BMI, coronary heart disease history, ...).

The impact will be significant if the method succeeds. Indeed, it will be possible for the nephrologists to use this model for monitoring individually each treatment, in order to avoid risks of hypotension (low BNP) or overweight (high BNP).

### 4.4. Intracellular processes

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.

- 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.

## 4.5. Population pharmacometrics

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.

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