RESEARCH CENTER

FIELD
Digital Health, Biology and Earth

Activity Report 2014

Section New Results

Edition: 2015-03-24

COMPUTATIONAL BIOLOGY

COMPUTATIONAL NEUROSCIENCE AND MEDECINE

EARTH, ENVIRONMENTAL AND ENERGY SCIENCES

MODELING AND CONTROL FOR LIFE SCIENCES

ABS Project-Team

5. New Results

5.1. Highlights of the Year

In 2014, two achievements are worth noticing:
Analysis of large assemblies using native mass spectrometry data. Native mass spectrometry is about to revolutionize structural biology, since such experiments give access to the composition in terms of subunits of large macro-molecular assemblies, usually beyond reach for classical experimental techniques. In this context, we designed an algorithm to infer pairwise contacts within subunits of large macro-molecular assemblies – see section 5.3.1 . To the best of our knowledge, our algorithm is the only one whose performances can be precisely analyzed, the contenders being of heuristic nature.

Analysis and comparison of conformational ensembles and sampled energy landscapes. A key property governing the behavior of many biophysical systems is the classical enthalpy - entropy balance, which is the root of thermodynamics. Therefore, studying the way a protein folds or the way two proteins assemble requires unveiling properties of ensembles of conformations of the system scrutinized. In this context, we designed novel methods to analyze and compare collections of conformations and the associated energy landscape – see section 5.4.1 . The algorithms are based on state-of-the-art techniques from computational topology (Morse theory, Morse homology), and optimal transportation.

5.2. Modeling Interfaces and Contacts

Docking, scoring, interfaces, protein complexes, Voronoi diagrams, arrangements of balls.

The work undertaken in this vein in 2014 will be finalized in 2015.

5.3. Modeling Macro-molecular Assemblies

Macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

5.3.1. Connectivity Inference in Mass Spectrometry based Structure Determination

Participants: Frédéric Cazals, Deepesh Agarwal.

In collaboration with C. Caillouet, and D. Coudert, from the COATI project-team (Inria - I3S (CNRS, University of Nice Sophia Antipolis)).

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists of finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting of finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts.

In this paper [14], we present two novel algorithms, MILP-W and MILP-WB. The former solves the minimum weight connectivity inference (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets.

Experiments on three systems (yeast exosome, yeast proteasome lid, human eiF3), for which reference contacts are known (crystal structure, cryo electron microscopy, cross-linking), show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work, typically a twofold increase in sensitivity.

The software accompanying this paper is made available, and should prove of ubiquitous interest whenever connectivity inference from oligomers is faced.

5.4. Modeling the Flexibility of Macro-molecules

Protein, flexibility, collective coordinate, conformational sampling dimensionality reduction.

5.4.1. Conformational Ensembles and Sampled Energy Landscapes: Analysis and Comparison

Participants: Frédéric Cazals, Tom Dreyfus, Christine Roth.

In collaboration with D. Mazauric (Inria Sophia Antipolis Méditerranée, Geometrica) and C. Robert (IBPC / CNRS, Paris).

In this work, we present novel algorithms and software addressing four core problems in computational structural biology, namely analyzing a conformational ensemble, comparing two conformational ensembles, analyzing a sampled energy landscape, and comparing two sampled energy landscapes [15]. Using recent developments in computational topology, graph theory, and combinatorial optimization, we make two notable contributions. First, we present a generic algorithm analyzing height fields. We then use this algorithm to perform density based clustering of conformations, and to analyze a sampled energy landscape in terms of basins and transitions between them. In both cases, topological persistence is used to manage (geometric) frustration. Second, we introduce two algorithms to compare transition graphs. The first is the classical *earth mover distance* metric which depends only on local minimum energy configurations along with their statistical weights, while the second incorporates topological constraints inherent to conformational transitions.

Illustrations are provided on a simplified protein model (BLN69), whose frustrated potential energy landscape has been thoroughly studied.

The software implementing our tools is also made available, and should prove valuable wherever conformational ensembles and energy landscapes are used.

5.5. Algorithmic Foundations

Computational geometry, Computational topology, Voronoi diagrams, α-shapes, Morse theory.

5.5.1. Mass Transportation Problems with Connectivity Constraints

Participant: Frédéric Cazals.

In collaboration with D. Mazauric (Inria Sophia Antipolis Méditerranée, Geometrica).

Given two graphs, the supply and the demand graphs, we analyze the mass transportation problem between their vertices, under connectivity constraints [16]. More precisely, for every subset of supply nodes inducing a connected component of the supply graph, we require that the set of demand nodes receiving non-zero flow from this subset induces a connected component of the demand graph. As opposed to the classical problem, a.k.a the earth mover distance (EMD), which is amenable to linear programming (LP), this new problem is very difficult to solve, and we make four contributions. First, we formally introduce two optimal transportation problems, namely minimum-cost flow under connectivity constraints problem (EMD-CC) and maximum-flow under cost and connectivity constraints problem (EMD-CCC). Second, we prove that the decision version of EMD-CC is NP-complete even for very simple classes of instances. We deduce that the decision version of EMD-CCC is NP-complete, and also prove that EMD-CC is not in APX even for simple classes of instances. Third, we develop a greedy heuristic algorithm returning admissible solutions, of time complexity $O(n^3m^2)$ with n and m the numbers of vertices of the supply and demand graphs, respectively. Finally, on the experimental side, we compare the transport plans computed by our greedy method against those produced by the aforementioned LP. Using synthetic landscapes (Voronoi landscapes), we show that our greedy algorithm is effective for graphs involving up to 1000 nodes. We also show the relevance of our algorithms to compare energy landscapes of biophysical systems (protein models).

5.5.2. *Ciruvis: a web-based Tool for Rule Networks and Interaction Detection using Rule-based Classifiers*

Participant: Simon Marillet.

In collaboration with J. Komorowski and S. Bornelöv (Uppsala University).

The use of classification algorithms is becoming increasingly important for the field of computational biology. However, not only the quality of the classification, but also its biological interpretation is important. This interpretation may be eased if interacting elements can be identified and visualized, something that requires appropriate tools and methods.

We developed a new approach to detecting interactions in complex systems based on classification [12]. Using rule-based classifiers, we previously proposed a rule network visualization strategy that may be applied as a heuristic for finding interactions. We now complement this work with Ciruvis, a web-based tool for the construction of rule networks from classifiers made of IF-THEN rules. Simulated and biological data served as an illustration of how the tool may be used to visualize and interpret classifiers. Furthermore, we used the rule networks to identify feature interactions, compared them to alternative methods, and computationally validated the findings. Rule networks enable a fast method for model visualization and provide an exploratory heuristic to interaction detection. The tool is made freely available on the web and may thus be used to aid and improve rule-based classification.

AMIB Project-Team

5. New Results

5.1. RNA

To mitigate the current absence of a selective scientific event dedicated to RNA computational biology, impeding the dissemination of recent methodological results, AMIB members have participated in the creation of the *Computational Methods for Structural RNAs* workshops (CMSR'14). This first installment of the event was hosted in Strasbourg as a workshop of the 2014 edition of European Conference on Computational Biology. Its proceedings were published by McGill University [33], and extended versions of selected articles were invited to appear in the *Journal of Computational Biology*.

5.1.1. RNA visualization

The field of RNA visualization is now rich with multiple tools that accommodate different needs, arising from a variety of application contexts. In order to help end-users navigate through the jungle of available options, Y. Ponty and F. Leclerc (IGM, Univ. Paris-Sud) have contributed a review of existing tools, and illustrate their usage to address a collection of typical use-cases [35].

5.1.2. RNA design and structures

The past couple of years have seen the multiplication of heuristic or exponential time algorithms for the RNA design problem. This situation motivates a survey, which s currently lacking, that would focus on the relative merits of existing algorithms, and assess their applicability towards the typical goals of synthetic biology. Such an objective evaluation is at the core of the PhD project of Vincent Le Gallic, which was started in September 2014.

With Antoine Soulé, a PhD student of J-M Steyaert and J. Waldispühl (McGill), a comparative study of the various softwares for the inverse RNA folding problem is under revision and a new version of RNAMUTANT in the langage GAP-L with enrichment has been designed.

Besides, we have published a general survey on RNA structure comparison [9].

5.1.3. RNA splicing regulation

RNA splicing is a modification of the nascent pre-messenger RNA (pre-mRNA) transcript in which introns are removed and exons are joined. The U2AF heterodimer protein has been well studied for its role in defining functional 3' splice sites in pre-mRNA splicing, but multiple critical problems are still outstanding, including the functional impact of their cancer-associated mutations. In collaboration with Xiang-Dong Fu's groups in San Diego and Wuhan, , through genome-wide analysis of U2AF-RNA interactions, we reported in [16] that U2AF has the capacity to define 88% of functional 3' splice sites in the human genome. Numerous U2AF binding events also occur in other genomic locations, and metagene and minigene analysis suggests that upstream intronic binding events interfere with the immediate downstream 3' splice site associated with either the alternative exon to cause exon skipping or competing constitutive exon to induce inclusion of the alternative exon.

5.1.4. RNA 3D structure modelling

Conformational diversity for RNA ensemble analyses is often provided by sophisticated molecular dynamics simulations. Long trajectories with specialized force fields on dedicated supercomputers are required to adequately sample conformational space, limiting ensemble analyses to modestly-sized RNA molecules. To avoid these limitations, we developed an efficient conformational sampling procedure, Kino-geometric sampling for RNA (KGSrna), which can report on ensembles of RNA molecular conformations orders of magnitude faster than MD simulations. In the KGSrna model, the RNA molecule is represented with rotatable, single bonds as degrees-of-freedom and groups of atoms as rigid bodies. In this representation, non-covalent bonds form distance constraints, which create nested, closed cycles in a rooted spanning tree. Torsional degrees-of-freedom in a closed ring demand carefully coordinated changes to avoid breaking the non-covalent bond, which greatly reduces the conformational flexibility. The reduced flexibility from a network of nested, closed rings consequently deforms the biomolecule along preferred directions on the conformational landscape. This new procedures projects degrees-of-freedom onto a lower-dimensional subspace of the conformation space, in which the geometries of the non-covalent bonds are maintained exactly under conformational perturbation. The dimensionality reduction additionally enables efficient exploration of conformational space and reduces the risk of overfitting sparse experimental data. Kinogeometric sampling of 3D RNA models can recover the conformational landscape encoded by proton chemical shifts in solution and is thus of great help to interpret NMR experimental data [11]. The computational efficiency of this approach, combined to its inherent parallel nature could also be adapted to model large assemblies on parallel platforms.

Our expertise was also essential in modelling junction of the RNA structure of a large biomolecule of interest, the tmRNA so as to study its interaction with the SmpB protein. Results obtained in collaboration with experiementalists, mainly P. Vachette at IBBMC and S. Nonin-Lecomte at the LCRB were made available in [15].

5.2. Sequences

5.2.1. Random generation

In collaboration with the Simon Fraser University (Vancouver, Canada), we have explored a random generation strategy, under a Boltzmann distribution, to assess the robustness of predicted adjacencies in ancestral genomes using a parsimony-based approach. The sampling algorithm was used to estimate the Boltzmann probability of ancestral adjacencies, which was then used as a filter to weed out unsupported predictions, leading to the resolution of a large number of syntenic inconsistencies [23].

5.2.2. Combinatorics of motifs

An algorithm for pvalue computation has been proposed in [40] that takes into account a Hiddden Markov Model and an implementation, SUFPREF, has been realized (http://server2.lpm.org.ru/bio).

Combinatorics of clumps have been extensively studied, leading to the definition of the so-called *canonic clumps*. It is shown in [26] that they contain the necessary information needed to calculate, approximate, and study probabilities of occurrences and asymptotics. This motivates the development of a *clump automaton*. It allows for a derivation of pvalues, decreasing the space and time complexity of the generating function approach or previous weighted automata. An extension to degenerate patterns is currently realized and implemented in a collaboration with J. Holub (Praha U.) and E. Furletova (IMPB).

During her master thesis at King's College, A. Héliou and collaborators designed the first linear-time and linear-space algorithm for computing all minimal absent words based on the suffix array [6]. In a typical application, one would be interested in computing minimal absent words to compare and study genomes in linear time by considering this negative information.

In a collaboration with AlFarabi University, where M. Régnier acts as a foreign co-advisor), word statistics were used to identify mRNA targets for miRNAs involved in various cancers [7].

5.2.3. Prediction and functional annotation of ortholog groups of proteins

In comparative genomics, orthologs are used to transfer annotation from genes already characterized to newly sequenced genomes. Many methods have been developed for finding orthologs in sets of genomes. However, the application of different methods on the same proteome set can lead to distinct orthology predictions.

In [38], [14] we developed a method based on a meta-approach that is able to combine the results of several methods for orthologous group prediction. The purpose of this method is to produce better quality results by using the overlapping results obtained from several individual orthologous gene prediction procedures. Our method proceeds in two steps. The first aims to construct seeds for groups of orthologous genes; these seeds correspond to the exact overlaps between the results of all or several methods. In the second step, these seed groups are expanded by using HMM profiles.

We evaluated our method on two standard reference benchmarks, OrthoBench and Orthology Benchmark Service. Our method presents a higher level of accurately predicted groups than the individual input methods of orthologous group prediction. Moreover, our method increases the number of annotated orthologous pairs without decreasing the annotation quality compared to twelve state-of-the-art methods.

5.3. 3D Modelling and Interactions

5.3.1. Transmembrane proteins

Transmembrane beta-barrel proteins (TMB) account for 20 to 30% of identified proteins in a genome but, due to difficulties with standard experimental techniques, they are only 2% of the RCSB Protein Data Bank. As TMB perform many vital functions, the prediction of their structure is a challenge for life sciences, while the small number of known structures prohibits knowledge-based methods for structure prediction. We study and design algorithmic solutions addressing the secondary structure, an abstraction of the 3D conformation of a molecule, that only retains the contacts between its residues. As TMBs are strongly structured objects, model based methodologies [18] are an interesting alternative to conventional methods. The efficiently obtained 3D structures provide a good model for further 3D and interaction analyses.

5.3.2. 3D Interaction prediction

While protein-RNA complexes provide a wide range of essential functions in the cell, their atomic experimental structure solving is even more difficult than for proteins. Protein-RNA complexes provide a wide range of essential functions in the cell. Docking approaches that have been developed for proteins are often challenging to adapt for RNA because of its inherent flexibility and the structural data available being relatively scarce.We adapted the reference RosettaDock protocol for protein-RNA complexes both at the nucleotide and atomic levels. Using a genetic algorithm-based strategy, and a non-redundant protein-RNA dataset, we derived a RosettaDock scoring scheme able not only to discriminate but also score efficiently docking decoys. The approach proved to be both efficient and robust for generating and identifying suitable structures when applied to two protein-RNA docking benchmarks in both bound and unbound settings. It also compares well to existing strategies. This is the first approach that currently offers a multi-level optimized scoring approach integrated in a full docking suite, leading the way to adaptive fully flexible strategies [28], [12]. This work is part of the PhD thesis of Adrien Guilhot-Gaudeffroy. While the previously described approaches perform well in a rigid or semi-flexible docking setting, the generation of putative conformations for flexible molecules (sampling) is still a difficult question that has to be addressed in a multi-scale setting involving new algorithms. Docking these sampled conformations will also certainly require improvement in clustering approaches.

5.4. Data Integration

With the increasing popularity of scientific workflows, public and private repositories are gaining importance as a means to share, find, and reuse such workflows.

As the sizes of these repositories grow, methods to compare the scientific workflows stored in them become a necessity, for instance, to allow duplicate detection or similarity search. Scientific workflows are complex objects, and their comparison entails a number of distinct steps from comparing atomic elements to comparison of the workflows as a whole. Various studies have implemented methods for scientific workflow comparison and came up with often contradicting conclusions upon which algorithms work best. Comparing these results is cumbersome, as the original studies mixed different approaches for different steps and used different evaluation data and metrics.

In collaboration with members of the University of Humboldt (Berlin), we first contribute to the field [17] by (i) comparing in isolation different approaches taken at each step of scientific workflow comparison, reporting on an number of unexpected findings, (ii) investigating how these can best be combined into aggregated measures, and (iii) making available a gold standard of over 2000 similarity ratings contributed by 15 workflow experts on a corpus of 1500 workflows and re-implementations of all methods we evaluated. In this context, we have designed new approaches based on consensus ranking [21] to provide a consensus of the experts' answers.

Then, with members of the University of Pennsylvania, we have presented a novel and intuitive workflow similarity measure that is based on layer decomposition [27] (designed during the month SCB spent at UPenn). Layer decomposition accounts for the directed dataflow underlying scientific workflows, a property which has not been adequately considered in previous methods. We comparatively evaluate our algorithm using our gold standard and show that it a) delivers the best results for similarity search, b) has a much lower runtime than other, often highly complex competitors in structure-aware workflow comparison, and c) can be stacked easily with even faster, structure-agnostic approaches to further reduce runtime while retaining result quality.

Another way to make scientific workflows easier to reuse is to reduce their structural complexity to make them easier to apprehend. In particular, we have continued to work in collaboration with the University of Manchester on DistillFlow, an approach to remove the structural redundancy in workflows. Our contribution is four fold. Firstly, we identify a set of anti-patterns that contribute to the structural workflow complexity. Secondly, we design a series of refactoring transformations to replace each anti-pattern by a new semantically-equivalent pattern with less redundancy and simplified structure. Thirdly, we introduce a distilling algorithm that takes in a workflow and produces a distilled semantically-equivalent workflow [8]. Lastly, we provide an implementation of our refactoring approach (dedicated demo published [24]) that we evaluate on both the public Taverna workflows and on a private collection of workflows from the BioVel project. On going work includes extending the list of anti-patterns to be considered and identifying *good patterns*, that is, patterns which are easy to maintain and have systematically been able to be executed. This has been done in the context of the master internship of Stéphanie Kamgnia Wonkap [37]. First results obtained are promising.

5.5. Systems Biology

5.5.1. Analyzing SBGN-AF Networks Using Normal Logic Programs

A wide variety of signaling networks are available in the literature or in databases under the form of influence graphs. In order to understand the systems underlying these networks and to modify them for a medical purpose, it is necessary to understand their dynamics. Consequently, a variety of modelling techniques for these networks have been developped. In particular, it is possible to model their dynamical behavior with Boolean networks. The construction of these Boolean networks starting from influence graphs requires a parametrization of some Boolean functions. This task is most often realized by interpreting experimental results, that can be hard to obtain.

We introduced a method that allows to model any influence graph expressed in the Systems Biology Graphical Notation Activity Flow language (SBGN-AF) under the form of a Boolean network [32], [29]. The parametrization does not rely on any experimental results but on general principles that govern the dynamics of signaling networks. Together with the translation of a SBGN-AF influence graph into predicates, these general principles expressed under the form of logic rules form a first-order normal logic program (NLP) equivalent to a Boolean network. We show that the trajectories as well as the steady-state of any SBGN-AF network can be obtained by computing the orbits and the supported models of its corresponding NLP, respectively.

5.5.2. Scalable methods for analysing dynamics of automata networks

In collaboration with T. Chatain, S. Haar, S. Schwoon, and L. Jezeguel (INRIA MEXICO), we explored new techniques for computing the reachable attractors in automata networks using Petri net unfoldings [22]. Attractors of network dynamics represent the long-term behaviours of the modelled system. Their characterization is therefore crucial for understanding the response and differentiation capabilities of a dynamical system. In the scope of qualitative models of interaction networks, the computation of attractors reachable from a given state of the network faces combinatorial issues due to the state space explosion. Our new algorithm relies on Petri net unfoldings that can be used to compute a compact representation of the dynamics, in particular by exploiting the concurrency of the transitions in order to remove redundant sequences of transitions. We illustrate the applicability of the algorithm with Petri net models of cell signalling and regulation networks, Boolean and multi-valued. The proposed approach aims at being complementary to existing methods for deriving the attractors of Boolean models, while being generic since it actually applies to any safe Petri net.

In collaboration with M. Folschette, M. Magnin, O. Roux (IRCCyN, Nantes), and K. Inoue (NII, Tokyo), we developed a framework for identifying classical Boolean or discrete networks models from Proces Hitting (PH) models [10]. The PH allows to model non-deterministic cooperations between interacting components, and we have shown that the dynamics of a single PH can embed (include) the dynamics of multiple discrete networks, where transitions functions are deterministic. Hence, if a behaviour is shown impossible at the PH model, it is necessary impossible in any included discrete models. Such kind of analysis is relevant in systems biology, where the cooperations between components are often under-determined and the enumeration of all compatible discrete models is intractable: our framework allows to reason on the dynamics of a single abstract model.

Finally, a chapter summarizing the recent advances on static analysis for dynamics of large biological networks has been published as part of the *Logical Modeling of Biological Systems* handbook [30].

BAMBOO Project-Team

6. New Results

6.1. Evolution of the genomes of endosymbionts in insects: the case of Hamiltonella defensa interacting with its various partners

Insect cells host many endosymbiotic bacteria, which are in general classified according to their importance for the host: “primary” symbionts are by definition mandatory and synthesize essential nutrients for the insects that feed on poor or unbalanced food sources, while “secondary” symbionts are optional and use mutualistic strategies and/or manipulation of reproduction to invade and persist within insect populations. *Hamiltonella defensa* is a secondary endosymbiont that established two distinct associations with phloemophagous insects. In aphids, it protects the host against parasitoid attacks. Its ability to infect many host tissues, notably the hemolymph, could promote its contact with parasitoid eggs. Despite this protective phenotype, the high costs associated with its presence within the host prevent its fixation in the population. In the whitefly *Bemisia tabaci* however, this symbiont is found only in cells specialised in hosting endosymbionts, the bacteriocytes. In these cells, it cohabits with other symbiotic species, such as the primary symbiont *Portiera aleyrodidarum*, a proximity that favors potential exchanges between the two symbionts. It is fixed in populations of *B. tabaci*, which suggests an important role for the consortium, probably nutritious.

We studied the specificities of each of these systems. First, in the bacteriocytes of *B. tabaci*, we identified a partitioning of the synthetic capacities of two endosymbionts, *H. defensa* and *P. aleyrodidarum*, in addition to a potential metabolic complementation between the symbionts and their host for the synthesis of essential amino acids. We proposed a key nutritive role for *H. defensa*, which would indicate a transition to a mandatory status in relation to the host and would explain its fixation in the population.

We also focused on the genomic evolution of the genus *Hamiltonella*, by comparing the strains infecting *B. tabaci* with a strain infecting the aphids. We highlighted the specialization of the symbionts to their hosts, and found that the genomes of the endosymbionts reflected their respective ecology. The aphid strain thus possesses many virulence factors and is associated with two partners, a bacteriophage and a recombination plasmid. These systems, inactive in the symbiont of *B. tabaci*, are directly related to the protection against and arms race with parasitoids. Conversely, the presumed avirulence of whitefly endosymbionts is consistent with their nutritional phenotype and a transition to a mandatory status to the host.

Finally, we studied the phenomenon of “accelerated mutation rate” in *H. defensa*, compared to its sister species *Regiella insecticola*, which is also a clade of protective endosymbionts of aphids. After excluding the assumption that the transition to the intracellular life occurred independently in the two lineages, we tried to establish a link between these differences in terms of evolvability in the endosymbionts and of their gene contents, particularly for genes involved in ecology and DNA repair. All the results obtained have provided insight into the evolution of the species *H. defensa*, since the last ancestor to the present species, by establishing a link between bacterial phenotype and genomic evolution.

The publications related to this area of research are either submitted or in preparation (to be submitted in the first months of the year).

6.2. Cardinium cBtQ1 providing insights into genome reduction, symbiont motility, and its settlement in Bemisia tabaci

Many insects harbor inherited bacterial endosymbionts. Although some of them are not strictly essential and are considered facultative, they can be a key to host survival under specific environmental conditions, such as parasitoid attacks, climate changes, or insecticide pressures. The whitefly *Bemisia tabaci* is at the top of the list of organisms inflicting agricultural damage and outbreaks, and changes in its distribution may be associated to global warming. In partnership with the group of Andrès Moya at the ICBiBE (Institut Cavanilles de Biodiversitat i Biologia Evolutiva), the genome of *Cardinium* cBtQ1, a facultative bacterial endosymbiont of *B. tabaci*, was sequenced and analysed [23].

6.3. Mitochondrial respiration and genomic analysis provide insight into the influence of the symbiotic bacterium on host trypanosomatid oxygen consumption

Certain trypanosomatids, such as *Angomonas deanei*, co-evolve with an endosymbiotic bacterium in a mutualistic relationship that is characterised by intense metabolic exchanges. We were able to show that the symbionts were able to respire for up to 4 h after isolation from the host. Moreover, our work suggests that the symbiont influences the mitochondrial respiration of the host protozoan [5].

6.4. Telling metabolic stories to explore metabolomics data

The increasing availability of metabolomics data enables to better understand the metabolic processes involved in the immediate response of an organism to environmental changes and stress. The data usually come in the form of a list of metabolites whose concentrations significantly changed under some conditions, and are thus not easy to interpret without being able to precisely visualize how such metabolites are interconnected.

We presented a method that enables to organize the data from any metabolomics experiment into metabolic stories [18]. Each story corresponds to a possible scenario explaining the flow of matter between the metabolites of interest. These scenarios may then be ranked in different ways depending on which interpretation one wishes to emphasize for the causal link between two affected metabolites: enzyme activation, enzyme inhibition or domino effect on the concentration changes of substrates and products. Equally probable stories under any selected ranking scheme can be further grouped into a single anthology that summarizes, in a unique subnetwork, all equivalently plausible alternative stories. An anthology is simply a union of such stories. We detailed an application of the method to the response of yeast to cadmium exposure. We used this system as a proof of concept for our method, and we showed that we are able to find a story that reproduces very well the current knowledge about the yeast response to cadmium. We further showed that this response is mostly based on enzyme activation. We also provided a framework for exploring the alternative pathways or side effects this local response is expected to have in the rest of the network. We discussed several interpretations for the changes we see, and we suggested hypotheses that could in principle be experimentally tested. Noticeably, our method requires simple input data and could be used in a wide variety of applications.

6.5. MiRNA and co: Methodologically exploring the world of small RNAs

We developed a reliable, robust, and much faster method for the prediction of pre-miRNAs. With this method, we aimed mainly at two goals: efficiency and flexibility. Efficiency was made possible by means of a quadratic algorithm. Since the majority of the predictors use a cubic algorithm to verify the pre-miRNA hairpin structure, they may take too long when the input is large. Flexibility relies on two aspects, the input type and the organism clade. MIRINHO can receive as input both a genome sequence and small RNA sequencing (sRNA-seq) data of both animal and plant species. To change from one clade to another, it suffices to change the lengths of the stem-arms and of the terminal loop. Concerning the prediction of plant miRNAs, because their pre-miRNAs are longer, the methods for extracting the hairpin secondary structure are not as accurate as for shorter sequences. With MIRINHO, we also addressed this problem, which enabled to provide premiRNA secondary structures more similar to the ones in MIRBASE than the other available methods.

Mirinho served as the basis to two other issues we addressed. The first issue led to the treatment and analysis of sRNA-seq data of *Acyrthosiphon pisum*, the pea aphid. The goal was to identify the miRNAs that are expressed during the four developmental stages of this species, allowing further biological conclusions concerning the regulatory system of such an organism. For this analysis, we developed a whole pipeline, called MIRINHOPIPE, at the end of which MIRINHO was aggregated.

We then moved on to the second issue, that involved problems related to the prediction and analysis of non-coding RNAs (ncRNAs) in the bacterium *Mycoplasma hyopneumoniae*. A method, called ALVINHO, was thus developed for the prediction of targets in this bacterium, together with a pipeline for the segmentation of a numerical sequence and detection of conservation among ncRNA sequences using a k-partite graph.

We finally addressed a problem related to motifs, that is to patterns, that may be composed of one or more parts, that appear conserved in a set of sequences and may correspond to functional elements. This had already been addressed in a robust method called Smile. However, depending on the input parameters, the output may be too large to be tractable, as was realized in other works of the team. We then presented some clustering solutions to group the motifs that may correspond to a same biological element, and thus to better distinguish the biologically significant ones from noise that may be present in what often are large outputs from many motif extraction algorithms.

The publications related to this area of research are either submitted or in preparation (to be submitted in the first months of the year).

6.6. Efficient Algorithms for analysing RNA-seq Data

In the last years, we had addressed the problem of identifying and quantifying variants (alternative splicing and genomic polymorphism) in RNA-seq data when no reference genome is available, without assembling the full transcripts. Based on the fundamental idea that each variant corresponds to a recognizable pattern, a bubble, in a de Bruijn graph constructed from the RNA-seq reads, we propose a general model for all variants in such graphs. We then introduced an exact algorithm, called KISSPLICE, to extract alternative splicing events. We had showed that it enables to identify more correct events than general purpose transcriptome assemblers.

The main time bottleneck in the KISSPLICE algorithm is the bubble enumeration step. Thus, in an effort to make our method as scalable as possible, we had modified Johnson's cycle listing algorithm (Johnson (1975)) to enumerate bubbles in general directed graphs, while maintaining the same time complexity. We now proposed, using a different enumeration technique, an algorithm to list bubbles with path length constraints in weighted directed graphs [29]. For a graph with n vertices and m edges, the method we propose lists all bubbles with a given source in $O(n(m + nlogn))$ delay. Moreover, we experimentally showed that this algorithm is several orders of magnitude faster than the listing algorithm of KISSPLICE to identify bubbles corresponding to alternative splicing events.

Additionally, we showed that the same techniques used to list bubbles can be applied to one classical enumeration problem: K-shortest paths problems [29]. We considered a different parameterisation of the K-shortest paths problem: instead of bounding the number of st-paths, we bound the weight of the st-paths. We present a general scheme to list bounded length st-paths in weighted graphs that takes $O(nt(n, m))$ time per path, where $t(n, m)$ is the time for a single source shortest path computation. This algorithm uses memory linear in the size of the graphs, independent of the number of paths output. For undirected non-negatively weighted graphs, we also show an improved algorithm that lists all st-paths with bounded length in $O((m + t(n, m)))$ time per path.

The main memory bottleneck in KISSPLICE is the construction and representation of the de Bruijn graph. Thus, again with the goal to make our method as scalable as possible, we propose a new compact way to build and represent a de Bruijn graph improving over the state of the art [22]. We show both theoretically and experimentally that our approach uses 30% to 40% less memory than such state of the art, with an insignificant impact on the construction time. Our de Bruijn graph representation is general, in other words it is not restricted to the variation finding or RNA-seq context, and can be used as part of any algorithm that represents NGS data with de Bruijn graphs.

A major issue when analysing transcriptomes using short sequencing reads is to be able to deal with repeats that are longer than the reads. We proposed a first explicit model for large families of inexact repeats in the de Bruijn Graphs generated from RNA-seq data [21]. Taking advantage of this modelling, we also proposed an efficient algorithm which enumerates alternative splicing events without traversing repeat-induced subgraphs, therefore offering a first answer to one the main question left open at the end of Gustavo Sacomoto's PhD [4].

Motivated by previous work on the classical problem of listing cycles, we also studied from a more purely theoretical point of view how to list chordless cycles [28]. We thus developed an amortized $\tilde{O}(|V|)$-delay algorithm for listing chordless cycles in undirected graphs. Chordless cycles are very natural structures in undirected graphs, with an important history and distinguished role in graph theory. The best known solution to list all the C chordless cycles contained in an undirected graph $G = (V, E)$ takes $O(|E|2 + |E|\cdot C)$ time. In this paper we provide an algorithm taking $\tilde{O}(|E| + |V|\cdot C)$ time. We also show how to obtain the same complexity for listing all the P chordless st-paths in G (where C is replaced by P).

6.7. Reference-free detection of isolated SNPs

Detecting single nucleotide polymorphisms (SNPs) between genomes is becoming a routine task with next-generation sequencing. Generally, SNP detec- tion methods use a reference genome. As non-model organisms are increasingly investigated, the need for reference-free methods has been amplified. Most of the existing reference-free methods have fundamental limitations: they can only call SNPs between exactly two datasets, and / or they require a prohibitive amount of computational resources. V. Lacroix participated in the developement of a method, called DISCOSNP to detect both heterozygous and homozygous isolated SNPs from any number of read datasets, without a reference genome, and with very low memory and time footprints (billions of reads can be analyzed with a standard desktop computer) [25]. To facilitate downstream genotyping analyses, DISCOSNP ranks predictions and outputs quality and coverage per allele. Compared to finding isolated SNPs using a state-of-the-art assembly and mapping approach, DISCOSNP requires significantly less computational resources, shows similar precision / recall values, and highly ranked predictions are less likely to be false positives. An experimental validation was conducted on an arthropod species (the tick *Ixodes ricinus*) on which de novo sequencing was performed. Among the predicted SNPs that were tested, 96% were successfully genotyped and truly exhibited polymorphism.

6.8. Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin

Alternative splicing is the main mechanism of increasing the proteome diversity coded by a limited number of genes. It is well established that different tissues or organs express different splicing variants. However, organs are composed of common major cell types, including fibroblasts, epithelial, and endothelial cells. By analysing large-scale data sets generated by The ENCODE Project Consortium and after extensive RT-PCR validation, we demonstrated that each of the three major cell types expresses a specific splicing program independently of its organ origin [17]. Furthermore, by analysing splicing factor expression across samples, publicly available splicing factor binding site data sets (CLIP-seq), and exon array data sets after splicing factor depletion, we identified several splicing factors that contribute to establishing these cell type-specific splicing programs.

6.9. Length and symmetry on the sorting by weighted inversions problem

Large-scale mutational events that occur when stretches of DNA sequence move throughout genomes are called genome rearrangement events. In bacteria, inversions are one of the most frequently observed rearrangements. In some bacterial families, inversions are biased in favour of symmetry as shown by recent research. In addition, several results suggest that short segment inversions are more frequent in the evolution of microbial genomes. Despite the fact that symmetry and length of the reversed segments seem very important, they have not been considered together in any problem in the genome rearrangement field. We defined the problem of sorting genomes (or permutations) using inversions whose costs are assigned based on their lengths and asymmetries [27]. We presented five procedures and assessed their performance on small sized permutations. The ideas presented in the paper provide insights to solve the problem and set the stage for a proper theoretical analysis.

6.10. Efficient tree reconciliation enumerator plus cophylogeny reconstruction algorithm via an Approximate Bayesian Computation

Phylogenetic tree reconciliation is the approach of choice for investigating the co-evolution of sets of organisms such as hosts and parasites. It consists in a mapping between the parasite tree and the host tree using event-based maximum parsimony. Given a cost model for the events, many optimal reconciliations are however possible. Any further biological interpretation of them must therefore take this into account, making the capacity to enumerate all optimal solutions a crucial point. Only two algorithms currently exist that attempt such enumeration; in one case not all possible solutions are produced while in the other not all cost vectors are currently handled. Our objective in addressing this problem was two-fold. The first was to fill this gap, and the second was to test whether the number of solutions generally observed can be an issue in terms of interpretation.

We presented a polynomial-delay algorithm called EUCALYPT for enumerating all optimal reconciliations [12]. We showed that in general many solutions exist. We gave an example where, for two pairs of host-parasite trees having each less than 41 leaves, the number of solutions is 5120, even when only time-feasible ones are kept. To facilitate their interpretation, those solutions were also classified in terms of how many of each event they contain. The number of different classes of solutions may thus be notably smaller than the number of solutions, yet they may remain high enough, in particular for the cases where losses have cost 0. In fact, depending on the cost vector, both numbers of solutions and of classes thereof may increase considerably (for the same instance, to respectively 4080384 and 275). To further deal with this problem, we introduced and analysed a restricted version where host-switches are allowed to happen only between species that are within some fixed distance along the host tree. This restriction allows us to reduce the number of time-feasible solutions while preserving the same optimal cost, as well as to find time-feasible solutions with a cost close to the optimal in the cases where no time-feasible solution is found.

Despite an increasingly vast literature on cophylogenetic reconstructions for studying host-parasite associations, understanding the common evolutionary history of such systems remains a problem that is far from being solved. Most algorithms for host-parasite reconciliation use an event-based model, where the events include in general (a subset of) cospeciation, duplication, loss, and host-switch. All known parsimonious event-based methods then assign a cost to each type of event in order to find a reconstruction of minimum cost. This is what we did ourselves in EUCALYPT. The main problem with this approach is that the cost of the events strongly influences the reconciliation obtained.

To deal with this problem, we developed an algorithm, called COALA, for estimating the frequency of the events based on an approximate Bayesian computation approach [8]. The benefits of this method are twofold: (1) it provides more confidence in the set of costs to be used in a reconciliation, and (2) it allows estimation of the frequency of the events in cases where the dataset consists of trees with a large number of taxa.

We evaluated our method on simulated and on biological datasets. We showed that in both cases, for the same pair of host and parasite trees, different sets of frequencies for the events lead to equally probable solutions. Moreover, often these solutions differ greatly in terms of the number of inferred events. It appears crucial to take this into account before attempting any further biological interpretation of such reconciliations. More generally, we also showed that the set of frequencies can vary widely depending on the input host and parasite trees. Indiscriminately applying a standard vector of costs may thus not be a good strategy.

6.11. Others

Other works, often experimental were also developed during 2014 and published in a number of papers [6], [7], [9], [10], [11], [13], [16], [19], [20], [24], [26].

BEAGLE Project-Team

5. New Results

5.1. Highlights of the Year

We organized two satellite workshops of international conferences:

- The Aevol tutorial during ALife 2014 (July 30 - August 2, New York) http://www.aevol.fr/alifeTutorial
- The "Computational Methods and Modeling of Astrocyte Physiology and Neuron-Glia Interactions" workshop during the Computational NeuroScience 2014 conference (July 26 - 31, Quebec City, Canada)

These highlight our active presence in the scientific life of our two sub-domains in major conferences.

5.2. Sparse short-distance connections enhance calcium wave propagation in a 3D model of astrocyte networks

Participants: H. Berry, J. Lallouette, M. De Pittá

Traditionally, astrocytes have been considered to couple via gap-junctions into a syncytium with only rudimentary spatial organization. However, this view is challenged by growing experimental evidence that astrocytes organize as a proper gap-junction mediated network with more complex region-dependent properties. On the other hand, the propagation range of intercellular calcium waves (ICW) within astrocyte populations is as well highly variable, depending on the brain region considered. This suggests that the variability of the topology of gap-junction couplings could play a role in the variability of the ICW propagation range. Since this hypothesis is very difficult to investigate with current experimental approaches, we explored it using a biophysically realistic model of three-dimensional astrocyte networks in which we varied the topology of the astrocyte network, while keeping intracellular properties and spatial cell distribution and density constant. Computer simulations of the model suggest that changing the topology of the network is indeed sufficient to reproduce the distinct ranges of ICW propagation reported experimentally. Unexpectedly, our simulations also predict that sparse connectivity and restriction of gap-junction couplings to short distances should favor propagation while long–distance or dense connectivity should impair it. Altogether, those results provide support to recent experimental findings that point towards a significant functional role of the organization of gap-junction couplings into proper astroglial networks. Dynamic control of this topology by neurons and signaling molecules could thus constitute a new type of regulation of neuron-glia and glia-glia interactions.

This result has been published in [18] and as conference talks. It is based on J. Lallouette's PhD thesis work in collaboration with M. De Pittà (postdoc in the team) and E Ben-Jacob, Tel Aviv University, Israel.

5.3. Glutamate Mediated Astrocytic Filtering of Neuronal Activity

Participants: H. Berry, J. Lallouette, M. De Pittá

Neuron-astrocyte communication is an important regulatory mechanism in various brain functions but its complexity and role are yet to be fully understood. In particular, the temporal pattern of astrocyte response to neuronal firing has not been fully characterized. Here, we used neuron-astrocyte cultures on multi-electrode arrays coupled to Ca2+ imaging and explored the range of neuronal stimulation frequencies while keeping constant the amount of stimulation. Our results reveal that astrocytes specifically respond to the frequency of neuronal stimulation by intracellular Ca2+ transients, with a clear onset of astrocytic activation at neuron firing rates around 3-5 Hz. The cell-to-cell heterogeneity of the astrocyte Ca2+ response was however large and increasing with stimulation frequency. Astrocytic activation by neurons was abolished with antagonists of type I metabotropic glutamate receptor, validating the glutamate-dependence of this neuron-to-astrocyte pathway. Using a realistic biophysical model of glutamate-based intracellular calcium signaling in astrocytes, we suggest that the stepwise response is due to the supralinear dynamics of intracellular IP3 and that the heterogeneity of the responses may be due to the heterogeneity of the astrocyte-to-astrocyte couplings via gap junction channels. Therefore our results present astrocyte intracellular Ca2+ activity as a nonlinear integrator of glutamate-dependent neuronal activity.

This result has been published in a paper currently in press, [26] and is a direct result from J. Lallouette's PhD thesis in collaboration with Y. Hanein's group, in Tel Aviv University (for the experimental measurements), M. De Pittà (postdoc in the team), and E Ben-Jacob, Tel Aviv University, Israel.

5.4. Space-induced bifurcation in repression-based transcriptional circuits

Participants: H. Berry, A. Lo Van

Experimental measurements of the mobility of macromolecules, especially proteins, in cells and their membranes consistently report transient subdiffusion with possibly position-dependent —non-homogeneous— properties. However, the spatiotemporal dynamics of protein mobility when transient subdiffusion is restricted to a subregion of space is still unclear. We have investigated the spatial distribution at equilibrium of proteins undergoing transient subdiffusion due to continuous-time random walks (CTRW) in a restricted subregion of a two-dimensional space. Our Monte-Carlo simulations suggest that this process leads to a non-homogeneous spatial distribution of the proteins at equilibrium, where proteins increasingly accumulate in the CTRW subregion as its anomalous properties are increasingly marked. These results suggest that, even though they exhibit the same time-dependence of the mean-squared displacement, the different scenarios proposed to account for subdiffusion in the cell lead to different protein distribution in space, even at equilibrium and without coupling with reaction. We also we assessed the influence of the spatial distribution of the genes on the dynamics of 3-gene transcriptional ring networks regulated by repression, i.e. repressilator circuits. Our simulations suggest that variations of spatial parameters – namely the degree of demixing of the positions of the gene or the spatial range of the mRNA and proteins (i.e. the typical distance they travel before degradation) – have dramatic effects by switching the dynamical regime from spontaneous oscillations to a stationary state where each species fluctuates around a constant value. By analogy with the bifurcations arising from the variation of kinetic parameters, we referred to those transitions as space-induced bifurcations. Therefore, our results strongly support the idea that the spatial organization of the molecular actors of transcriptional networks is crucial for the dynamics of gene expression and suggest that the spatial localization of the synthetic genes in the cell could be used as an additional toggle to control the dynamics of the inserted construct in synthetic biology experiments.

This group of results has been published in [20], [13], [12] and [23]. It consists in the PhD and Master works of B. Caré and A. Lo Van, respectively, and a collaboration with H Chaté, CEA, Saclay.

5.5. Modeling interaction of transcription processes in neighbour genes

Participants: G. Beslon, S. Meyer

During the transcription process, the genetic sequence encoded in the DNA molecule is expressed by an enzymatic complex. This process is often considered as independent for each gene, despite numerous reported cases of one transcribed gene perturbing a neighbour gene's expression, which is then regarded as a side-effect. Here, we suggest in the contrary that such interactions are a widespread feature, resulting from the propagation along the DNA molecule of mechanical stress generated during gene transcription. This torsional stress modifies the facility with which the transcription machinery separates the two strands of the double-helix in order to access the bases, and thus the expression level of any gene located nearby. We develop a quantitative model of this effect, showing that it depends strongly on the orientation of the genes, which is confirmed by the analysis of in vivo expression levels in the drosophila genome. This observation suggests that torsional coupling may play an important role in genetic regulation, and might favor the orientation-dependent co-localization of genes involved in similar functions, which need to be expressed together.

Publication: [21]

5.6. A model of genome size evolution

Participants: G. Beslon, C. Knibbe, S. Fisher

Even though numerous genomic sequences are now available, evolutionary mechanisms that determine genome size, notably their fraction of non-coding DNA, are still debated. In particular, although several mechanisms responsible for genome growth (proliferation of transposable elements, gene duplication and divergence, etc.) were clearly identified, mechanisms limiting the overall genome size remain unclear.

In collaboration with Samuel Bernard (Inria Dracula Team and Institut Camille Jordan, UMR CNRS 5208, Lyon), we have developed a model for genome size evolution that takes into account both local mutations such as small insertions and small deletions, and large chromosomal rearrangements such as duplications and large deletions. We introduced the possibility of undergoing several mutations within one generation. The model, albeit minimalist, revealed a non-trivial spontaneous dynamics of genome size: in the absence of selection, an arbitrary large part of genomes remains beneath a finite size, even for a duplication rate 2.6-fold higher than the rate of large deletions, and even if there is also a systematic bias toward small insertions compared to small deletions. Specifically, we showed that the condition of existence of an asymptotic stationary distribution for genome size non-trivially depends on the rates and mean sizes of the different mutation types. We also gave upper bounds for the median and other quantiles of the genome size distribution, and argue that these bounds cannot be overcome by selection. Taken together, these results show that the spontaneous dynamics of genome size naturally prevents it from growing infinitely, even in cases where intuition would suggest an infinite growth. This work was part of Stephan Fischer's PhD thesis, which was defended in December 2013.

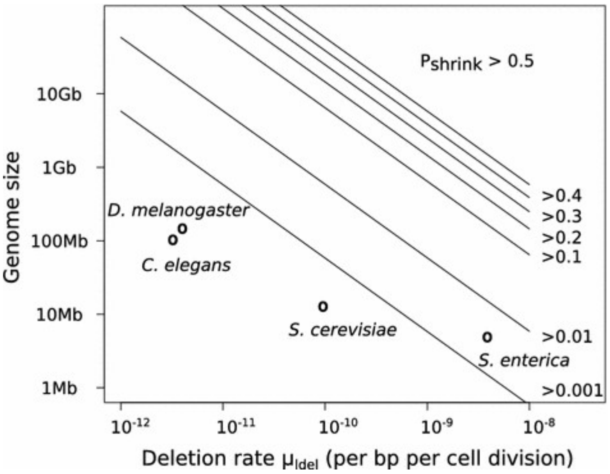


Figure 2. Comparison of the bounds on genome size with the genome size for four organisms. Spontaneous deletion rates were computed per base pair and per cell division from experimental data on mutation accumulations for the bacterium Salmonella enterica, the budding yeast Saccharomyces cerevisiae, the worm Caenorhabditis elegans and the fruit fly Drosophila melanogaster. The value next to each line is the lower bound for the probability that a genome located along this line will shrink at the next step in our model for equal duplication and deletion rates.

This year, using quantitative numerical examples with parameters taken from biological data, we showed that, in practice, a shrinkage bias appears very quickly in genomes undergoing mutation accumulation, even though DNA gains and losses appear to be perfectly symmetrical at first sight. This spontaneous dynamics provides the genome with a stability-related size limit below which it can be influenced by other evolutionary forces (selection, drift, biases, ...).

All this work has been published this year [15], and is already mentioned as "most read article" by Springer.

5.7. A novel view on reductive evolution

Participants: G. Beslon, C. Knibbe, B. Batut

Bacterial genomes show substantial variations in size. The smallest bacterial genomes are those of endocellular symbionts of eukaryotic hosts, which have undergone massive genome reduction and show patterns that are consistent with the degenerative processes that are predicted to occur in species with small effective population sizes. However, similar genome reduction is found in some free-living marine cyanobacteria that are characterized by extremely large populations. Using a combination of bioinformatics approaches and of in silico experimental evolution (with the aevol model), we have been able to propose a scenario that explains the reductive evolution of marine bacteria.

This work was part of Bérénice Batut's PhD thesis [10], which was defended in November 2014. Bérénice was co-supervised by Guillaume Beslon and Carole Knibbe (Inria BEAGLE team) for the simulations and by Gabriel Marais and Vincent Daubin (Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558) for the genomic analyses. This work had already yielded a publication in 2013 [34]. This year, we published a review in the high-level journal [11]. The scenario proposed in the PhD manuscript, as well as the simulations and analyses done this year to support it, should be published in 2015.

5.8. Genome evolution aware gene trees

Participant: E. Tannier

Traditionally the inference of a gene tree is made from a multiple alignment of homologous sequences according to a model of molecular evolution. Trees for several gene families are thus constructed one by one, independently from each other. Constructed this way trees often carry unresolutions or bad resolutions. Information for their full resolution may lie in the poorly exploited dependency between gene families, each bringing information for the resolution of the others. We used several kinds of such dependencies in the construction of gene trees: information from a species tree through a model of gene content evolution, information from extant synteny through ortholog predictions, and information from ancestral synteny through a model of gene neighborhood evolution. We developed, improved, implemented and gave a user interface to several "correction" techniques, yielding a series of correction modules called "RefineTree". We tested its parts on simulated data and apply it on the full set of gene families from the Ensembl Compara database. We showed that according to several measures including the tree likelihood computed from sequence evolution, the stability of genome content and the linearity of ancestral chromosomes, trees corrected by refineTree are arguably more plausible than the ones stored by Ensembl.

This work has been achieved by Magali Semeria, Laurent Gueguen (LBBE) and Eric Tannier in Lyon, in collaboration with Nadia El-Mabrouk's group from the computer science department of the university of Montreal. This collaboration started when Nadia El-Mabrouk was an Inria visiting professor in our team in 2012 and 2013. An article has been submitted.

5.9. Variable food availability increases weight: a mathematical prediction

Participant: H. Soula

Due to the conservation of energy, the energy storage in adipose tissue reflect the difference of energy expenditure and energy intake. Without change in physical activity, the main paradigm has always been that this storage does not depend on the timing of intake but on its whole temporal integration: the overall food intake. However, mammal and especially rats can compensate energy expenditure to save energy in case of starving. This adaptation should provoke variation in energy expenditure when food availability varies in time. Using animal experiments and mathematical modelling, we showed that indeed food availability variation - while conserving the same amount of energy - can disrupt and perturb energy balance. Submitted to variation in availability with a period above 4 weeks, rats where bigger with higher fat mass than control. Even so these rats had eaten the same amount of food as the control group during the same period. Our mathematical model uses delay equations and can predict both the food intake and the body weight variations. We showed that delay in energy saving adaptation cause this variation and estimate the lag at 1 week. This result could very well apply to humans in the so called ’yoyo regime’. Regime that are stopped are a typical case of food intake variation and could cause greater fat accretion instead of body weight reduction. We show that this should happen if the regime lasts longer than one week.

This result has been the subject of an article in the weekly journal of Inserm Rhônes-Alpes with an interview of author H. Soula.

5.10. Insights on gene family dynamics from digital genetics experiments

Participants: C. Knibbe

Gene families are sets of homologous genes formed by duplications of a single original gene. Inferring their history in terms of gene duplications, gene losses and gene mutations yields fundamental insights into the molecular basis of evolution. However, the traditional approach, the phylogenetic inference of gene family evolution, faces two difficulties: (i) the delimitation of gene families based on sequence similarity, and (ii) the fact that the models of evolution used for reconstruction are tested against simulated data that are produced by the model itself. This year, we showed that digital genetics, or in silico experimental evolution, can provide thought-provoking synthetic gene family data, robust to rearrangements in gene sequences and, most importantly, not biased by where and how we think natural selection should act. Using Aevol, we analyzed the evolution of 3,512 synthetic gene families under directional selection. The turnover of gene families in evolutionary runs was such that only 21% of those families would be accessible for classical phylogenetic inference. Extinct families showed patterns different from the final, observable ones, both in terms of dynamics of gene gains and losses and in terms of gene sequence evolution. This study also reveals that gene sequence evolution, and thus evolutionary innovation, occurred not only through local mutations, but also through chromosomal rearrangements that re-assembled parts of existing genes.

This work was published in the international conference ALIFE 2014 [28].

BIGS Project-Team

5. New Results

5.1. Analysis of high dimensional data

Participants: K. Duarte, S. Ferrigno, J.-M. Monnez, A. Muller-Gueudin, S. Tindel

5.1.1. Online partial principal component analysis of a data stream

Consider a data stream and suppose that each data vector is a realization of a random vector whose expectation varies with time, the law of the centered data vector being stationary. Consider the principal component analysis (PCA) of this centered vector called partial PCA. In this study are defined online estimations of the first principal axes by stochastic approximation processes using a data batch at each step of the process or all the data until the current step. This extends a former result obtained by J.-M. Monnez by using one data vector at each step. This is applied to partial generalized canonical correlation analysis by defining a stochastic approximation process of the metric involved in this case using all the data until the current step. If the expectation of the data vector varies according to a linear model, a stochastic approximation process of the model parameters is used. All these processes can be performed in parallel. A forthcoming preprint by R. Bar and J.-M. Monnez will discuss those aspects.

5.1.2. Data analysis for cumulative exposure Index

Everyone is subject to environmental exposures from various sources, with negative health impacts (air, water and soil contamination, noise ...) or with positive effects (e.g., green space). Studies considering such complex environmental settings in a global manner are rare. In [5] we propose to use statistical factor and cluster analyses to create a composite exposure index with a data-driven approach, in view to assess the environmental burden experienced by populations. The study was carried out in the Great Lyon area (France, 1.2M inhabitants) at the census block group (BG) scale. We used as environmental indicators ambient air NO2 annual concentrations, noise levels, proximity to green spaces, to industrial plants, to polluted sites and to road traffic. Although it cannot be applied directly for risk or health effect assessment, the resulting index can help to identify hot spots of cumulative exposure, to prioritize urban policies or to compare the environmental burden across study areas in an epidemiological framework.

5.1.3. A simultaneous stepwise covariate selection

In supervised learning the number of values of a response variable to predict can be high. Also clustering them in a few clusters can be useful to perform relevant supervised classification analysis. On the other hand selecting relevant covariates is a crucial step to build robust and efficient prediction models, especially when too many covariates are available in regard to the overall sample size. As a first attempt to solve these problems, we had already devised in a previous study an algorithm that simultaneously clusters the levels of a categorical response variable in a limited number of clusters and selects forward the best covariates by alternate minimization of Wilks's Lambda. In the project carried out this year, we first extend the former version of the algorithm to a more general framework where Wilks's Lambda can be replaced by any model selection criterion. We also turned forward selection into stepwise selection in order to remove covariates in real time if necessary. Finally an application of our algorithm to real datasets from peanut allergy studies allowed to get confirmation of some previously published results and suggested new discoveries. The possibilities of this algorithm are promising and it is hoped to be useful for many practitioners.

5.1.4. Prognostic value of the Strauss estimated plasma

We describe here an application oriented study lead jointly by J.-M. Monnez and a medical team under the supervision of E. Albuisson at CHU Brabois. The objective is to assess the prognostic value of estimations of volemia, or of their variations, beyond clinical examination in a post-hoc analysis of the Eplerenone Post-Acute Myocardial Infarction (AMI) Heart Failure (HF) Efficacy and Survival Study (EPHESUS). Assessing congestion post-discharge is indeed challenging but of paramount importance to optimize patient management and prevent hospital readmissions. The analysis was performed in a subset on 4957 patients with available data (within a full dataset of 6632 patients). Study endpoint was cardiovascular death and/or hospitalization for HF between month 1 and month 3 after post-AMI HF. Estimated plasma volume variations between baseline and month 1 were estimated by the Strauss formula, which includes hemoglobin and hematocrit ratios. Other potential predictors including congestion surrogates, hemodynamic and renal variables, and medical history variables were tested. An instantaneous estimation of plasma volume at month 1, ePVS M1, was defined and also tested. Multivariate analysis was performed using stepwise logistic regression and linear discriminant analysis. In HF complicating MI, congestion assessed by the Strauss formula and an instantaneous derived measurement of plasma volume displayed an added predictive value of early cardiovascular events, beyond routine clinical assessment. Trials assessing congestion management guided by this simple tool to monitor plasma volume are warranted.

5.1.5. Non parametric estimation of the conditional cumulative distribution function

This project fits into the global aim of improving local regression techniques. Indeed, we propose in [21] to study the local linear estimator of the conditional distribution function. Namely, having an i.i.d. sample $(X_i, Y_i)_{1\leq i\leq n}$, we estimate the conditional distribution function $F(t|x) = \mathbb{P}(Y \leq t|X = x)$ by:

$$\widehat{F}_n^{(1)}(t,h_n|x) = \frac{\widehat{f}_{n,2}(x,h_n)\widehat{r}_{n,0}(x,t,h_n) - \widehat{f}_{n,1}(x,h_n)\widehat{r}_{n,1}(x,t,h_n)}{\widehat{f}_{n,0}(x,h_n)\widehat{f}_{n,2}(x,h_n) - \left(\widehat{f}_{n,1}(x,h_n)\right)^2} \tag{1}$$

where ${}^{(1)}$ denotes the order 1 of the local polynomial estimator, $\widehat{f}_{n,j}$ stands for a kernel estimator with order j of the probability density function f_X of X, $\widehat{r}_{n,j}$ estimates the distribution of the couple (X,Y) and h_n is a bandwidth parameter.

This estimator is a particular case of the local polynomial estimators. It is the local polynomial estimator of order $p = 1$. Another simpler estimator, with order $p = 0$, is well known as the Nadaraya-Watson estimator.

We are interested in showing the advantage of this estimator over the Nadaraya-Watson estimator. We show asymptotic results for our estimator (exact rate of uniform consistency), and establish also uniform asymptotic certainty bands for the conditional cumulative distribution function.

We obtain the following result under some assumptions on the cumulative distribution F, f_X, the kernel K and the bandwidth h_n,

$$\sup_{t\in\mathbb{R}}\sup_{x\in I}\sqrt{\frac{nh_n}{\log(h_n^{-1})}}\left|\widehat{F}_n^{(1)}(t,h_n|x) - \widehat{\mathbb{E}}\left(\widehat{F}_n^{(1)}(t,h_n|x)\right)\right| \xrightarrow[n\to+\infty]{\mathbb{P}} \sigma_F(I) \tag{2}$$

where

$$\sigma_F^2(I) = \frac{||K||_2^2}{2\inf\limits_{x\in I} f_X(x)}.$$

As corollaries of this result, we extend our results to other statistical functions, such as the quantiles and the regression function.

We illustrate our results with simulations and an application on foetopathologic data.

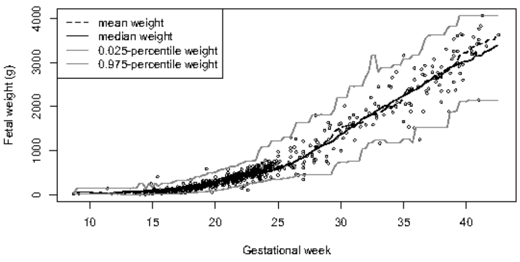


Figure 5. Fetal weight during the pregnancy: estimation of mean and quantiles from our local polynomial regression method.

We have also started a study about the regression function in the application on foetopathologic data. We consider the nonparametric model

$$Y = m(X) + \epsilon,$$

where Y is the fetal weight , X are the gestational weeks, m is a smooth unknown function and ϵ the error. The goal is to provide a test to detect significant features (or change points) of this regression curve. The regression curve is estimated using local polynomial kernel smoothers.

5.2. Stochastic modeling for complex and biological systems

Participants: R. Azaïs, T. Bastogne, C. Lacaux, A. Muller-Gueudin, S. Tindel, P. Vallois, S. Wantz-Mézières

5.2.1. Modelisation of networks of multiagent systems

We relate here the beginning of collaboration between A. Gueudin, R. Azaïs and some automatic control researchers in Nancy.

We consider networks, modeled as a graph with nodes and edges representing the agents and their interconnections, respectively. The connectivity of the network, persistence of links and interactions reciprocity influence the convergence speed towards a consensus.

The problem of consensus or synchronization is motivated by different applications as communication networks, power and transport grids, decentralized computing networks, and social or biological networks.

We then consider networks of interconnected dynamical systems, called agents, that are partitioned into several clusters. Most of the agents can only update their state in a continuous way using only inner-cluster agent states. On top of this, few agents also have the peculiarity to rarely update their states in a discrete way by resetting it using states from agents outside their clusters. In social networks, the opinion of each individual evolves by taking into account the opinions of the members belonging to its community. Nevertheless, one or several individuals can change its opinion by interacting with individuals outside its community. These inter-cluster interactions can be seen as resets of the opinions. This leads us to a network dynamics that is expressed in term of reset systems. We suppose that the reset instants arrive stochastically following a Poisson renewal process.

5.2.2. Tumor growth modeling

A cancer tumor can be represented for simplicity as an aggregate of cancer cells, each cell behaving according to the same discrete model and independently of the others. Therefore to measure its size evolution, it seems natural to use tools coming from dynamics of population, for instance the logistic model. This deterministic framework is well-known but the stochastic one is worthy of interest. We are currently studying in [22] a model in which we suppose that the size V_t at time t of the tumor is a diffusion process of the type :

$$\begin{cases} dV_t = r\,V_t\,\left(1 - \frac{V_t}{\kappa}\right) - c\,V_t + \beta\,V_t\,dB_t \\ V_0 = v > 0 \end{cases} \tag{3}$$

where $(B_t)_{t\geq 0}$ is a standard brownian motion starting from zero.Then (i) We define a family of time continuous Markov chains which models the evolution of the rate of malignant cells and approximate (under some conditions) the diffusion process (V_t). (ii) We study in depth the solution to equation (3). This diffusion process lives in a domain delimited by two boundaries: 0 and $\kappa > 0$. In this stochastic setting, the role of κ is not so clear and we contribute to understand it. We describe the asymptotic behavior of the diffusion according to the values of the parameters. The tools we resort to are boundary classification criteria and Laplace transform of the hitting time to biological worthwhile level. We are able in particular to express the mean of the hitting time.

5.2.3. Anisotropic random fields

Hermine Biermé (Tours) and Céline Lacaux follow in [19] their collaboration in the study of anisotropic random fields. They have extended their previous works in the framework of conditionally sub-Gaussian random series. For such anisotropic fields, they have obtained a modulus of continuity and a rate of uniform convergence. Their framework allows to study e.g., Gaussian fields, stable random fields and multi-stable random fields.

5.2.4. Inference for dynamical systems driven by Gaussian noises.

As mentioned in the *Scientific Foundations* Section, the problem of estimating the coefficients of a general differential equation driven by a Gaussian process is still largely unsolved. To be more specific, the most general ($\mathbb{R}$-valued) equation handled up to now as far as parameter estimation is concerned is of the form:

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

where θ is the unknown parameter, b is a smooth enough coefficient and B is a one-dimensional fractional Brownian motion. In contrast with this simple situation, our applications of interest (see the *Application Domains* Section) require the analysis of the following $\mathbb{R}^n$-valued equation:

$$X_t^\theta = a + \int_0^t b(\theta; X_u)\,du + \int_0^t \sigma(\theta; X_u)\,dB_t, \tag{4}$$

where θ enters non linearly in the coefficient, where σ is a non-trivial diffusion term and B is a d-dimensional fractional Brownian motion. We have thus decided to tackle this important scientific challenge first.

To this aim, here are the steps we have focused on in 2014:

- A better understanding of the underlying rough path structure for equation (4). This includes two studies on differential systems driven by some general Gaussian noises in infinite dimensions: [17] on the Parabolic Anderson model, and [16] about viscosity solutions in the rough paths setting.
- Study of densities for general systems driven by Gaussian noises as in [18] and [15].
- Ergodic aspects, which are another important ingredient for estimation procedures for stochastic differential equations, are handled in [3].

5.2.5. Extremal process

In extreme value theory, one of the major topics is the study of the limiting behavior of the partial maxima of a stationary sequence. When this sequence is i.i.d., the unique limiting process is well-known and called the extremal process. Considering a long memory stable sequence, the limiting process is obtained as a simple power time change extremal process. Céline Lacaux and Gennady Samorodnistky have proved in [23] that this limiting process can also be interpreted as a restriction of a self-affine random sup measure. In addition, they have established that this random measure arises as a limit of the partial maxima of the same long memory stable sequence, but in a different space. Their results open the way to propose new self-similar processes with stationary max-increments.

5.2.6. Self-nested structure of plants

In a recent work, Godin and Ferraro designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

The degree of self-nestedness is defined as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus design a heuristic method based on interacting simulated annealing algorithms to tackle this difficult question. This procedure is also a keystone in a new topological clustering algorithm for trees that we propose in this work. In addition, we obtain new theoretical results on the combinatorics of self-nested structures. For instance, we have shown that the number $C_{\leq H}(m)$ of self-nested trees with maximal height H and a ramification number for each vertex less than m satisfies the following formula,

$$C_{\leq H}(m) = \sum_{h=1}^{H} \prod_{i=1}^{h} \binom{m+h-i}{h-i+1}.$$

In particular, the cardinality $C_{=h}(m)$ of self-nested trees with exact height h evolves according to

$$\log C_{=h}(m) \sim \frac{(m+h)^2}{2} \log (m+h) - \frac{h^2}{2} \log h - \frac{m^2}{2} \log m - mh \log m,$$

when m and h simultaneously go to infinity. The redaction of an article is currently in progress.

5.2.7. Semi-parametric inference for a growth-fragmentation model

Statistical inference for piecewise-deterministic Markov processes has been extensively investigated for a few years under some ergodicity conditions. Our paper [2] is dedicated to a statistical approach for a particular non ergodic growth-fragmentation model for which the set $[0, 1]$ is absorbing. This kind of stochastic process may model the dynamic of a malthusian population for which there exists an extinction threshold. We focus on the estimation of the extinction probability and of the distribution of the extinction time from only one path of the model within a long time interval.

We establish that the absorption probability p is the unique solution in an appropriate space of a Fredholm equation of the second kind whose parameters are unknown,

$$p - Kp = s,$$

where K is an integral operator depending explicitly on the main features of the model. From n data, we estimate this important characteristic of the underlying process by solving numerically the estimated Fredholm equation. Indeed, $\widehat{p}_{n,m}$ is defined as the approximated solution of the equation $p - \widehat{K}_n p = \widehat{s}_n$ after m steps of the algorithm. Fortunately, this procedure allows us to estimate also the extinction time.

We have shown the convergence in probability of the proposed estimators under some usual asymptotic conditions. In particular, we have,

$$\forall\, \varepsilon > 0,\ \mathbf{P}\left(\|p - \widehat{p}_{n,m}\|_1 > \varepsilon\right) \to 0,$$

when n and m simultaneously go to infinity. The good behavior of our estimates on finite sample sizes is presented in Figure 6 . In future works, we plan to apply this methodology to more intricate situations, in particular for the pharmacokinetics and pharmacodynamics stochastic models recently introduced in the literature.

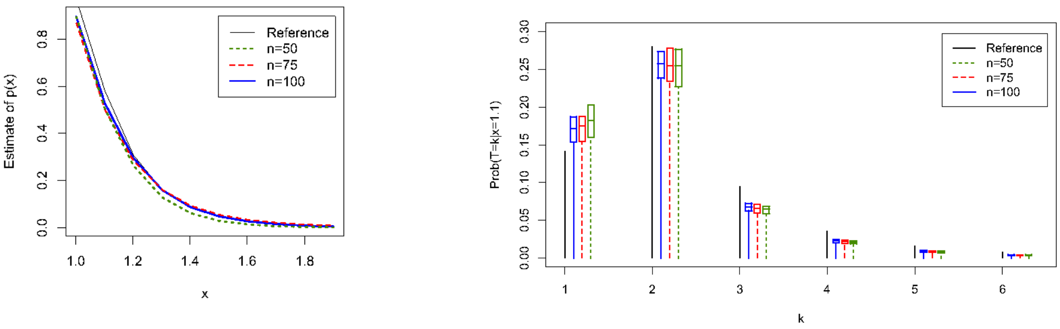


Figure 6. Probability of extinction $p(x)$ and its estimates for a trajectory starting from the initial population $1 \leq x \leq 2$ (left) and distribution of the extinction time and its estimates for a trajectory starting from $x = 1.1$ (right).

5.2.8. *A Model-based Pharmacokinetics Characterization Method of Engineered Nanoparticles for Pilot Studies*

Recent developments on engineered multifunctional nanomaterials have opened new perspectives in oncology. But assessment of both quality and safety in nanomedicine requires new methods for their biological characterization. We have recently proposed a new model-based approach for the pre-characterization of multifunctional nanomaterials pharmacokinetics in small scale in vivo studies. Two multifunctional nanoparticles, with and without active targeting, designed for photodynamic therapy guided by magnetic resonance imaging are used to exemplify the presented method. It allows the experimenter to rapidly test and select the most relevant pharmacokinetic (PK in the sequel) model structure planned to be used in the subsequent explanatory studies. We also show that the model parameters estimated from the in vivo responses provide relevant preliminary information about the tumor uptake, the elimination rate and the residual storage. For some parameters, the accuracy of the estimates is good enough to compare and draw significant pre-conclusions. A third advantage of this approach is the possibility to optimally refine the in vivo protocol for the subsequent explanatory and confirmatory studies complying with the 3Rs (reduction, refinement, replacement) ethical recommendations. More precisely, we show that the identified model may be used to select the appropriate duration of the magnetic resonance imaging sessions planned for the subsequent studies. The proposed methodology integrates magnetic resonance image processing, continuous-time system identification algorithms and statistical analysis. Except, the choice of the model parameters to be compared and interpreted, most of the processing procedure may be automated to speed up the PK characterization process at an early stage of experimentation.

More specifically, our efforts have been split into the following tasks:

- The article [6] gives an application of statistical methods for the design of experiments to optimize the formulation of a composite molecule in photodynamic therapy. The associated know-how has been transferred to the start-up CYBERnano to be generalized to the rational design of engineered nanoparticles. Collaboration with CRAN and LRGP (Nancy) and UNINE (Neuchâtel, Suisse).
- In [12], in vivo application of photodynamic therapy, a mathematical model and computational simulations of the light propagation in biological tissues were developed to help biologists to determine *a priori* some parameters of the experimental protocol. More precisely, the numerical results were used to select the most suited position of the optical fiber to be implemented within the animal brain. This equipment is required to bring the light and thus activate the molecule within the tumor. The therapeutical objective was to maximize the homogeneity of light intensity within the tumor volume.
- Obstacles and challenges to the clinical use of the photodynamic therapy (PDT) are numerous: large inter-individual variability, heterogeneity of therapeutic predictability, lack of in vivo monitoring concerning the reactive oxygen species (ROS) production, etc. All of these factors affect in their ways the therapeutic response of the treatment and can lead to a wild uncertainty on its efficiency. To deal with these variability sources, we have designed and developed an innovative technology able to adapt in realtime the width of light impulses during the photodynamic therapy. The first objective is to accurately control the photobleaching trajectory of the photosensitizer during the treatment with a subsequent goal to improve the efficacy and reproducibility of this therapy. In this approach, the physician a priori defines the expected trajectory to be tracked by the photosensitizer photobleaching during the treatment. The photobleaching state of the PS is regularly measured during the treatment session and is used to change in real-time the illumination signal. This adaptive scheme of the photodynamic therapy has been implemented, tested and validated during in vitro tests. These tests show that controlling the photobleaching trajectory is possible, confirming the technical feasibility of such an approach to deal with inter-individual variabilities in PDT. These results, contained in [13], open new perspectives since the illumination signal can be different from a patient to another according to his individual response. This study has proven its interest by showing promising results in an in vitro context, which has to be confirmed by the current in vivo experiments. However, it is fair to say that in a near future, the proposed solution could lead, in fine, to an optimized and personalized PDT. A patent was deposited subsequently. Collaboration with CRAN (Nancy).
- The communications [8], [9] and [10] present successful applications of a model-based design of nanoparticles. This approach is based on statistical design of experiments and black-box modeling in cell biology. The associated know-how has been transferred to the start-up CYBERnano. Collaboration with CEA LETI and INSERM (Grenoble).

BONSAI Project-Team

6. New Results

6.1. Highlights of the Year

- Amandine Perrin received the best paper award and the best oral presentation at the ISCB-LA 2014 international conference for the work on reconstruction of ancestral gene orders.
- Hélène Touzet was invited as a keynote speaker at the ALGO 2014 international conference. The topic of the talk was RNA bioinformatics.

BEST PAPERS AWARDS :
[7] **ISCB-Latin America**. A. PERRIN, J.-S. VARRÉ, S. BLANQUART, A. OUANGRAOUA.

6.2. High-throughtput sequence processing

- **Analysis of immunological rearrangements for leukemia diagnosis and monitoring.** High-throughput sequencing is spreading in the hospitals and many classical routines are now being transferred to this new technology. However in the specific case of lymphocyte monitoring, some complications arise. Classical bioinformatics software tools do not apply to the specificity of lymphocyte rearrangements. That is why we developed the software Vidjil (see 5.2) together with Lille hospital. This work has been published [5] and was also presented, as a poster, during the annual conference of the American Society of Hematology (ASH) [13]. We are now members of the EuroClonality-NGS work group which aims at providing a standardized way of monitoring leukemia using high-throughput sequencing at the European level.
- **New seeds for approximate pattern matching.** We addressed the problem of approximate pattern matching using the Levenshtein distance. Given a text T and a pattern P, find all locations in T that differ by at most k errors from P. For that purpose, we proposed a filtration algorithm that is based on a novel type of seeds, combining exact parts and parts with a fixed number of errors, that we called 01^*0 seed. Implementation has been performed on a Burrows-Wheeler transform. Experimental tests show that the method is specifically well-suited to search for short patterns (< 50 letters) on a small alphabet (*e.g.* DNA alphabet) with a medium to high error-rate (7 %–15 %). This work has been published in [9], and has a large number of applications in computational biology, such as finding microRNA targets, for example.
- **Spaced seeds and Transition seeds.** This year, two collaborative works have been published on the topic of spaced seeds and derivated models. The first work, resulting from a collaboration with Martin C. Frith from the *Computational Biology Research Center* (Tokyo), increases the sensivity of several search tools (among them, LAST, LASTZ, YASS,...) by computing specific seeds adapted to transition ratios observed during Eucaryotic comparisons. This work has been published in [3], together with the design of seeds obtained. The second work, issued from collaboration with Donald E.K. Martin from the *Department of Statistics* of the *North Carolina State University* (Raleigh), deals with the coverage of spaced seeds and shows how this criterion helps selecting good seeds for SVM string-kernels and alignment-free distances. This work has been published in [6].

6.3. RNA algorithms

- **A universal framework for RNA algorithms.** We have proposed a new generic specification framework, called *inverted coupled rewrite systems* that can deal with optimization problems on strings, trees, and arc-annotated sequences. It is specifically well-suited to handle RNA algorithms, such as alignment or folding algorithms. It is based on the following ideas. The solutions of combinatorial optimization problems are the inverse image of a term rewrite relation that reduces problem solutions to problem inputs. A tree grammar is used to further refine the search space, and optimization objectives are specified as interpretations of these terms. All these constituents provide a mathematically precise and complete problem specification, leading to concise yet translucent specifications of dynamic programming algorithms. This work is a collaborative project with R. Giegerich from Universität Bielefeld, and has been published in [4].

- **RNA multistructures.** In many RNA families, the signature of the family cannot be characterized by a single consensus structure, and is mainly described by a set of alternate secondary structures. For example, certain classes of RNAs adopt at least two distinct stable folding states to carry out their function. This is the case of riboswitches, that undergo structural changes upon binding with other molecules, and recently some other RNA regulators were proven to show evolutionary evidence for alternative structure. The necessity to take into account multiple structures also arises when modeling an RNA family with some structural variation across species, or when it comes to work with a set of predicted suboptimal foldings. In this perspective, we have introduced the concept of RNA multistructures, that is a formal grammar based framework specifically designed to model a set of alternate RNA secondary structures. We provide several motivating examples and propose an efficient algorithm to search for RNA multistructures within a genomic sequence. This work was published in [8].

6.4. Ancestral gene order reconstruction

- In the field of **genomic rearrangement**, a topic of interest is to infer ancestral gene order from gene order known in extant species. The problem resumes to compute a set ancestral CARs (continuous ancestral regions) at a given node of a phylogeny. We designed a progressive homology-based method which iteratively detects and assembles ancestral adjacencies while allowing some micro-rearrangements of synteny blocks at the extremities of the progressively assembled CARs. Comparing with other methods we are able to produce more robust CARs with a very simple and efficient method. This work was published in [7].

6.5. Nonribosomal peptides

- **Monomeric structure.** The algorithm that identifies the monomeric structure of a polymer from its chemical structure has been finished and named s2m. It is based on a double index: A partial index constructed on the monomer database that uses a markovian model to speed up the search time ; and an index constructed on the fly on the studied polymer. This strategy was originally developed for nonribosomal peptides, but can be applied to any polymer.
- **Florine: Nonribosomal peptide synthetase annotations.** Florine [2] is a workflow dedicated to the discovery of new nonribosomal peptide synthetases. It describes sequential steps starting from DNA sequences leading to the design of candidate bioactive peptides. It is a useful tool for new drug discovery because it can be applied whatever the producing micro-organisms as it takes into account the enzymatic specificities related to each genus. This work was performed in collaboration with members of EPI Orpailleur (CRI Nancy Grand Est), Marie-Dominique Devignes and Malika Smaïl-Tabbone.
- **Activity prediction of small molecules.** Bayesian Belief Network was used for the first time to classify compounds according to their biological activity [1]. This method was applied on nonribosomal peptides and gave promising results on predicting their activity.

DYLISS Project-Team

6. New Results

6.1. Highlights of the Year

Four PhD theses were defended this year. They evidenced that ASP-technologies are now mature enough to perform data integration of large-scale bio-molecular datasets: classification of families of proteins [10], reconstruction of regulatory networks [13], reconstruction of metabolic network [11], and modelling of the discrete dynamics of a signalling or a regulatory network [12]. Importantly, symbolic classification technics have been adapted to exhibit relevant biological features: we used both formal concept analysis and semantic-based analysis for sequence and network analysis.

6.2. Data integration

Participants: Jacques Nicolas, Charles Bettembourg, Jérémie Bourdon, Jeanne Cambefort, Marie Chevallier, Guillaume Collet, Olivier Dameron, Damien Eveillard, Julie Laniau, Sylvain Prigent, Anne Siegel, Valentin Wucher.

Pan-genomic metabolic network of *Ectocarpus siliculosus*: We introduced the first metabolic network for the non-classical species *E. Siliculosus*, called EctoGEM. The reconstruction process includes draft reconstruction based on sequence and functional annotation analysis. It is followed by a combinatorial gap-filling process using the Meneco software based on answer set programming, a semantic analysis of the completion and a manual curation. This reconstruction enables a better understanding of organism biology and a reannotation of its genome. [*J. Cambefort, G. Collet, O. Dameron, D. Eveillard, S. Prigent, A. Siegel*] [22], [11]

New insights on bacteria associated with brown algae As an application of our tools for the reconstruction of metabolic networks, we have contributed to the analysis of the genome of a bacteria which lives in symbiosis with brown algae by investigating candidates for metabolic exchanges between the bacteria and the algae. [*G. Collet, J. Cambefort, A. Siegel*] [19] [Online publication]

Modeling parsimonious putative regulatory networks We integrated heterogeneous information from two types of network predictions to determine a causal explanation for the observed gene co-expression. We modeled this integration as a combinatorial optimization problem. We demonstrated that this problem belongs to the NP-hard complexity class. We proposed an heuristic approach to have an approximate solution in a practical execution time. Our evaluation showed that the E.coli regulatory network resulting from the application of this method has higher accuracy than the putative one built with traditional tools. [*A. Aravena, A. Siegel, D. Eveillard*] [26] [Online publication]

Modeling of a gene network between mRNAs and miRNAs to predict gene functions involved in phenotypic plasticity in the pea aphid and non coding RNA in pea aphid During its PhD, V. Wucher has built the first network model of gene regulation by microRNAs in pea aphid. The thesis has studied the discrimination between embryos development towards either sexual or asexual reproduction types in the pea aphid *Acyrthosiphon pisum*, at the genomic level. The study of the post-transcriptional regulation network implies both the identification of regulated elements during embryogenesis and the identification of the interaction modules between microRNAs and mRNAs using formal concept analysis. It helps the understanding of regulation differences between sexual and asexual embryogenesis. Moreover, it is the first step towards the modeling of the entire set of genes regulations in pea aphid during embryogenesis. [*V. Wucher, J. Nicolas, F. Legeai (Genscale team)*] [13], [39], [30]

Using a large-scale knowledge database on reactions and regulations to exhibit key regulators A new formalism of regulated reactions combining biochemical transformations and regulatory effects was proposed to unify the different mechanisms contained in knowledge libraries. Based on a related causality graph, an algorithm was developed to propose a reasonable set of upstream regulators from lists of target molecules. Scores were added to candidates according to their ability to explain the greatest number of targets or only few specific ones. The method was validated on a real example related to glycolysis. [*P. Blavy, A. Siegel*] [18] [Online publication]

Semantic particularity measure for functional characterization of gene sets using gene ontology We propose a new approach to compute gene set particularities based on the information conveyed by Gene Ontology terms. A GO term informativeness can be computed using either its information content based on the term frequency in a corpus, or a function of the term's distance to the root. We demonstrated that the combination of semantic similarity and semantic particularity measures was able to identify genes with particular functions from among similar genes. This differentiation was not recognized using only a semantic similarity measure. [*C. Bettembourg, O. Dameron*] [17] [Online publication]

Integrating GALAXY workflows in a metadata management environment New tools are needed to enable the quick design and the intensive parallel execution of bioinformatics processes. Therefore, we proposed a new dataflow-oriented workflow management system dedicated to intensive bioinformatics tasks. We worked on the interoperatibility of bioinformatics workflows using a model-driven approach. Our results enable new import / export capabilities beetwen multiple workflow management environnements and insights to create a unique shared workflow model. [*O. Dameron, F. Moreews (Genscale team), Y. Le Bras (GenOuest platform), C. Monjeaud (GenOuest platform), O. Collin (GenOuest platform)*][36]

6.3. Time-series and asymptotic dynamics

Participants: Anne Siegel, Jérémie Bourdon, Jeanne Cambefort, Damien Eveillard, Vincent Picard, Nathalie Théret, Santiago Videla.

Reasoning on the response of logical signaling networks with boolean models A series of papers and a PhD thesis focused on modeling the response of logical signaling networks by means of automated reasoning using ASP. In this context, a crucial issue is automatic learning of logical networks from partial observations of input/output behaviours, in order to achieve unbiased and robust discoveries. Experiments showed that many networks can be compatible with a given set of experimental observations. In a review chapter, we first discuss how ASP can be used to exhaustively enumerate all these logical networks. Next, in order to gain control over the system, we look for intervention strategies that force a set of target species into a desired steady state. Finally, we discuss the usage of ASP for solving the aforementioned problems and the novelty of our approach with respect to existing methods. [*S. Videla, A. Siegel, J. Nicolas*] [23], [38], [12] [Online publication]

Integrative modeling framework for signaling networks based on guarded transitions models We develop a new non-ambiguous formal interpretation of signaling pathways as discrete dynamic models. The resulting language, Computer-Aided Design for BIOlogical Models (CADBIOM), is based on a simplified version of guarded transitions in which we introduced temporal parameters for each transition to manage competition and cooperation between parts of the models . Tools for simulation and model checking analyses using the formal Cadbiom language have been developed (http://cadbiom.genouest.org). Using CADBIOM, we built the first discrete model of TGF-β signaling networks by automatically integrating the 137 human signaling maps from the Pathway Interaction Database into a single unified dynamic model. Temporal property-checking analyses of 15934 trajectories that regulate 145 TGF-$beta$ target genes reveal the association of specific pathways with distinct biological processes. [*G Andrieux, M Le Borgne, N. Théret*] [15] [Online publication]

Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs We extended a Flux-Balanced-Analysis approach to quantify the precursor composition of each system output and to discuss the biological relevance of a set of flux in a given metabolic network. The composition is called contribution of inputs over outputs [AIO]. In order to further investigate metabolic network flexibility, we have proposed an efficient local search algorithm computing the extremal values of

AIO coefficients. This approach enables to discriminate diets without making any assumption on the internal behaviour of the system. [*O. Abdou-Arbi, J. Bourdon, A. Siegel*] [14] [Online publication]

Multivariate Normal Approximation for the Stochastic Simulation Algorithm: limit theorem and applications We prove a central limit theorem on the asymptotic stochastic dynamical behavior of the outputs of a reaction network under certain steady-state assumptions. We present multiple applications including a constraints-based approach to verify asymptotic properties on the output moments without prior knowledge about the kinetic parameters. [*V. Picard, A. Siegel, J. Bourdon*] [33] [Online publication]

Numeric model for initiation of translation in sea-urchin We use a numeric-based modeling approach to study the regulation of protein synthesis following fertilization in sea urchin. This approach based on parcimonious modelling evidenced that two processes are required to explain experimental data: a destabilization of eIF4E:4E-BP complex and a great stimulation of the 4E-BP-degradation mechanism, both rapamycin-sensitive [*A. Siegel, J. Bourdon*] [21] [Online publication]

6.4. Sequence annotation

Participants: François Coste, Aymeric Antoine-Lorquin, Catherine Belleannée, Guillaume Collet, Gaëlle Garet, Clovis Galiez, Laurent Miclet, Jacques Nicolas, Valentin Wucher.

Automated Enzyme Classification by Formal Concept Analysis Guessing enzyme's functional activity from its sequence is a crucial task that can be approached by comparing the new sequences with those of already known enzymes labeled by a family class. This task is difficult because the activity is based on a combination of small sequence patterns and sequences greatly evolved over time. We have designed a classifier based on the identification of common subsequence blocks between known and new enzymes and the search of formal concepts built on the cross product of blocks and sequences for each class. Since new enzyme families may emerge, it is important to propose simultaneously a first classification of enzymes that cannot be assigned to a known family. Formal Concept Analysis offers a nice framework to set this task as an optimization problem on the set of concepts. The classifier has been tested with success on a particular set of enzymes present in a large variety of species, the haloacid dehalogenase (HAD) superfamily. [*F. Coste, G. Garet, J. Nicolas*] [28], [10]

A bottom-up efficient algorithm learning substitutable languages from positive examples Based on Harris's substitutability criterion, the recent definitions of classes of substitutable languages have led to interesting polynomial learnability results for expressive formal languages. These classes are also promising for practical applications: in natural language analysis, because definitions have strong linguisitic support, but also in biology for modeling protein families, as suggested in our previous study introducing the class of local substitutable languages. But turning recent theoretical advances into practice badly needs truly operable algorithms. We present here an efficient learning algorithm, motivated by intelligibility and parsing efficiency of the result, which directly reduces the positive sample into a small non redundant canonical grammar of the target substitutable language. Thanks to this new algorithm, we have been able to extend our experimentation to a complete protein dataset confirming that it is possible to learn grammars on proteins with high specificity and good sensitivity by a generalization based on local substitutability. [*F. Coste, G. Garet, J. Nicolas*] [29], [10]

Logol: Expressive Pattern Matching in sequences. Application to Ribosomal Frameshift Modeling Logol consists in both a language for describing biological patterns, and an associated parser for effective pattern search in sequences (RNA, DNA or protein). The Logol language, based on an high level grammatical formalism (String Variable Grammars), allows to express flexible patterns (with mispairings and indels) composed of both sequential elements (such as motifs) and structural elements (such as repeats or pseudoknots). Its expressive power allows the design of sophisticated patterns such as the signature of "-1 programmed ribosomal frameshifting" (PRF) events in messenger RNA sequences. A PRF signature is a complex model composed of a slippery site followed by a pseudoknot located in a specific part of the sequence, which provides a good illustration of the Logol language power. [*C. Belleannée, J. Nicolas, O. Sallou (GenOuest platform)*] [27] [Online publication]

Identifying distant homologous viral sequences in metagenomes using protein structure information It is estimated that marine viruses daily kill about 20% of the ocean biomass. Identifying them in water samples is thus a biological issue of great importance. The metagenomic approach for virus identification is a challenging task since their sequences carry a lot of mutations and are very difficuly to identify by standard homology searches. The PEPS VAG project aims at establishing a novel methodology that uses structures of proteins as extra-information in order to annotate metagenomes without relying on sequence homology. In the context of the first experiments made on the metagenome of station 23 of the TARA Ocean Project, we used the structures of capsid proteins to infer the sequence signature of their fold, in order to find them in the metagenome. This work presents the methodology, the first experiments and the on-going improvements. [*C. Galiez, F. Coste*] [35]

Computational Protein Design: trying an Answer Set Programming approach to solve the problem The problem of *Computational Protein Design* aims at finding the best protein conformation to perform a given task. This problem can be reduced to an optimization problem, looking for the minimum of an energy function depending on the amino-acid interactions in the protein. The CPD problem may be easily modeled as an ASP program but a practical implementation able to work on real-sized instances has never been published. We have raised the main source of difficulty for current ASP solvers and ran a series of benchmarks highlighting the importance of finding a good upper bound estimation of the target minimum energy to reduce the amount of combinatorial search. Our solution clearly outperforms a direct ASP implementation without this estimation and has comparable performances with respect to SAT-based approaches. It remains less efficient than a recent approach by cost function networks, showing there still exists some place for improving the optimization component in ASP with more dynamical strategies. [*J. Nicolas, H. Bazille*] [34]

Searching for Optimal Orders for Discretized Distance Geometry The Molecular Distance Geometry Problem (MDGP) is the problem of finding the possible conformations of a molecule by exploiting available information on some distances between pairs of its atoms. When some assumptions are satisfied, the MDGP can be discretized, so that the search domain of the problem becomes a tree where each node corresponds to a candidate position for an atom. The search tree can be efficiently explored by using an *interval* Branch & Prune (iBP) algorithm that can potentially enumerate all feasible conformations. In this context, the order given to the atoms of the molecule plays an important role, because it allows the discretization assumptions to be satisfied, and it also impacts the computational cost of the iBP algorithm. We have proposed a new discretization order for protein backbones based on the optimization of certain criteria for a faster exploration of the discretized search domain. To this aim, we express the search for optimal orders by a set of logical constraints in ASP. Our comparison with previously proposed orders for protein backbones shows that this new discretization order makes iBP perform better. [*J. Nicolas, A. Mucherino (Genscale Team)*] [43]

From analogical proportions in lattices to proportional analogies in formal concepts We provided an attempt at bridging formal concept analysis and the modeling of analogical proportions (i.e., statements of the form “a is to b as c is to d”). A suitable definition for analogical proportions in non distributive lattices is proposed and then applied to concept lattices. This enables us to compute what we call proportional analogies. In addition, we define the locally maximal subwords and locally minimal superwords common to a finite set of words. We also define the corresponding sets of alignments. We show that the constructed family of sets of alignments has the lattice structure. The study of analogical proportion in lattices gives hints to use this structure as a machine learning basis, aiming at inducing a generalization of the set of words. [*L. Miclet*] [32], [37]

GENSCALE Project-Team

6. New Results

6.1. Highlights of the Year

discoSnp published in NAR. The publication presents a wide range of discoSnp applications that highlight the advantages and the drawbacks of predicting SNPs when no reference genomes are available. The publication witnesses the enthusiasm of users regarding both reference-free methods and the quality of the method. [20]

6.2. NGS methodology

Participants: Erwan Drezen, Anaïs Gouin, Dominique Lavenier, Claire Lemaitre, Antoine Limasset, Pierre Peterlongo, Guillaume Rizk.

Comparison of large sets of metagenomics data

We significantly extend the previous method (implemented in the Comparead tool) for computing similarity between sets of raw non assembled (and usually non-assemblable with current state of the art assemblers) reads. This enhancement of the method enables computations to be factorized when N read sets have to be compared all together. Moreover, the great advantage of this improvment is to save huge disk space and to enable efficient logical operations between metagenomic subset of reads. The Commet tool implements this optimized version.[25]

De novo SNP discovery

We developed a very efficient new way for detecting isolated SNPs given one, two or more raw read set(s) without using any reference genome. The implementation, called discoSnp, was applied to various datasets and applications. In particular, compared to finding isolated SNPs using a state-of-the-art assembly and mapping approach, our method requires significantly less computational resources, shows similar precision/recall values, and highly ranked predictions are less likely to be false positives. An experimental validation was conducted on an arthropod species (the tick *Ixodes ricinus*) on which de novo sequencing was performed. Among the predicted SNPs that were tested, 96% were successfully genotyped and truly exhibited polymorphism. [20]

De novo discovery of inversion breakpoints

A formal model has been proposed, together with an algorithm, for detecting inversion breakpoints without a reference genome, directly from raw NGS data. This model is characterized by a fixed size topological pattern in the de Bruijn Graph. We describe precisely the possible sources of false positives and false negatives and we additionally propose a sequence-based filter giving a good trade-off between precision and recall of the method. We implemented these ideas in a software called TakeABreak. Applied on simulated inversions in genomes of various complexity (from E. coli to a human chromosome dataset), the method provided promising results with a low memory footprint and a small computational time. [24]

Integrated detection and assembly of long insertion variants

We investigated a new method for the integrated detection and assembly of insertion variants from re-sequencing data. Contrary to other tools, it is designed to call insertions of any size, whether they are novel or duplicated, homozygous or heterozygous in the donor genome. We uses an efficient k-mer based method to detect insertion sites in a reference genome, and subsequently assemble them from the complete set of donor reads. The method is implemented in the tool MindTheGap and showed high recall and precision on simulated datasets of various genome complexities. When applied to real *C. elegans* and human datasets, MindTheGap detected and correctly assembled insertions longer than 1 kb, using at most 14 GB of memory. [19], [40]

Enhancement of de-Bruijn Graph data structure

The data structure holding the de-Bruijn Graph at the core of the GATB library has been improved through several new developments. First, its construction time has been greatly decreased thanks to the use of minimizers for kmer-counting, and efficient parallelization of various construction steps. Secondly, exploration of the graph has also been made faster through the possibility of parallel enumeration of nodes of interest, and through the use of a cache-coherent (blocked) bloom filter. Lastly, the structure itself has been extended to optionally allow for more information to be held, at a reasonable memory cost. A minimal perfect hash function allows to store additional data for each node, for example the coverage of each kmer. [11], [35], [36]

Chloroplast assembly

When sequencing plants, reads that correspond to the chloroplast genome are often over-represented. Filtering these reads based on k-mer counts allows specific assembly of the chloroplast to be directly performed. The small number of contigs can then be processed using advanced optimization tools to generate scaffolds. The approach has been partially tested on seqencing data from *Lactococcus lactis* to assemble plasmids of this bacteria. [12]

6.3. NGS applications

Participants: Susete Alves Carvalho, Rumen Andonov, Anaïs Gouin, Fabrice Legeai, Dominique Lavenier, Claire Lemaitre, Pierre Peterlongo, Ivaylo Petrov, Guillaume Rizk.

Identification of genomic regions of biological interest

The extraction and selection of 400 microsatellites among the large and fragmented *Acyrthosiphon pisum* genome led to the identification of a single 9cM region controlling the loss of sex in the pea aphid. The genotyping of these markers on geographically distant populations under divergent selection for reproductive strategies revealed a strong signature of selection in this genomic region, suggesting gene flow between populations with distinct reproductive modes. [15]

Transcriptome assembly

For this study, we incorporated various sources of RNA sequences from 454, Illumina and Sanger sequencing technologies obtained from more than 35 *S. frugiperda* developmental time-points and tissue samples and developed a custom pipeline to achieve their assembly. As a result, we provided a first valid transcriptome for *Spodoptera frugiperda*, a major agricultural pest. [16]

Catalogue of long non coding RNAs

We established a new bioinformatics pipeline for the detection of lncRNAs from RNA-Seq data, produced the first catalogue of aphids lncRNAs, and asserted for each lncRNA a classification of putative cis-interactions based on its genomic distance to neighboring mRNAs. These results allow the constitution of a broad gene regulation network of the aphid phenotypic plasticity at the embryo level. This workflow is available in Galaxy on the BioInformatics Platform for Arthopods of Agroecosystems (www.inra.fr/bipaa) and can be applied to any organism for which an annotated genome sequence and RNA-Seq data are provided.[23]

Identification and correction of genome mis-assemblies due to heterozygosity

Assembly tools are more and more efficient to reconstruct a genome from next-generation sequencing data but some problems remain. One of them corresponds to mis-assemblies due to heterozygosity (2 alleles instead of a consensus). Thus, we propose a strategy to detect and correct false duplications in assemblies based on several metrics: sequence similarity, matche length and average read coverage. Our method allows to decrease redundancy in the genome assembly, to improve the scaffolding and then to increase the N50 statistic by removal of one of the two alleles or joining of scaffolds by their extremities. This method was applied on the *Spodoptera frugiperda* genome.[39]

Questioning the classical re-sequencing analyses approach

Classical re-sequencing analyses are based on a first step of read mapping, then only mapped reads are taken into account in following analyses such as variant calling. We investigated the sources of unmapped reads in aphid re-sequencing data of 33 individuals, and we demonstrated that these reads contain valuable information that should not be discarded as usually done in such analyses. For instance, the analysis of the contigs obtained from assembling the unmapped reads led to recover some divergent genomic regions previously excluded from analysis and to discover putative novel sequences of *A. pisum* and its symbionts. We proposed strategies, based on assembly and re-mapping, to aid the capture and interpretation of this information.[14]

Application of discoSnp on pea data
The pea is a non-model organism with a large (4.5 GB) and complex genome which has not been sequenced yet. We compared, on the same set of low depth pea sequences, the SNPs generated by discoSnp with those obtained with a previous SNP discovery pipeline, and those generated using classical mapping approach combining Bowtie2 and GATK tools. [31]

6.4. HPC and parallelism

Participants: Rumen Andonov, Charles Deltel, Dominique Lavenier, François Moreews, Ivaylo Petrov.

Workflows
New tools are needed to enable the quick design and the intensive parallel execution of bioinformatics processes. Therefore, we propose a new Dataflow oriented workflow management system dedicated to intensive bioinformatics tasks. We worked on the interoperability of bioinformatics workflows using a model driven approach. Our results enable new import / export capabilities between multiple workflow management environments and incite to create a unique shared workflow model.[28]

Graph processing : the All-Pairs Shortest Paths problem
This research work anticipates the need of processing huge graphs that are results of intensive genomic sequence comparison (bank to bank processing). We proposed a new algorithm for solving the all-pairs shortest-path problem for planar graphs and graphs with small separators that exploits the massive on-chip parallelism available in today's Graphics Processing Units (GPUs). Our algorithm, based on the Floyd-War shall algorithm, has near optimal complexity in terms of the total number of operations, while its matrix-based structure is regular enough to allow for efficient parallel implementation on the GPUs. By applying a divide-and-conquer approach, we are able to make use of multi-node GPU clusters, resulting in more than an order of magnitude speedup over the fastest known Dijkstra-based GPU implementation and a two-fold speedup over a parallel Dijkstra-based CPU implementation.[27]

Benchmark of Alignment Search Tools
Comparing sequences is a daily task in bioinformatics and many software try to fulfill this need by proposing fast execution times and accurate results. Introducing a new software in this field requires to compare it to recognized tools with the help of well defined metrics. A set of quality metrics is proposed that enables a systematic approach for comparing alignment tools. These metrics have been implemented in a dedicated software, allowing to produce textual and graphical benchmark artifacts. [21]

6.5. Protein Structure

Participants: Rumen Andonov, Douglas Goncalves, Dominique Lavenier, Mathilde Le Boudic-Jamin, Antonio Mucherino.

The molecular distance geometry problem
The distance geometry is the problem of finding an embedding of a simple weighted undirected graph $G = (V, E, d)$ in a given dimension $K > 0$. Its most interesting application arises in biology, where the conformation of molecules such as proteins can be identified by embedding a graph (representing the molecular structure and some distance information) in dimension 3. Since some years, we are working on the discretization of the distance geometry. This year, the research developed in 4 main directions, that will be briefly detailed in the following paragraphs.

The majority of the work was performed on the so-called *discretization orders*, which are particular orders for the atoms of a molecule that allow for satisfying the discretization assumptions, i.e. they allow to discretize the search domain of the problem. Finding discretization orders is therefore an important pre-processing step for the solution of distance geometry problems. In fact, not only the identification of an atomic order allowing for the discretization is important, but also the identification of orders that are able to optimize some objectives that make the solution to the problem easier to perform. In this context, with both international and local partners, we worked on discretization orders that can be identified automatically in polynomial time [13], we worked on suitable orders for the protein side chains [10], and we studied some objectives to be optimized in discretization orders [38].

The algorithm that we mostly employ for the solution of distance geometry problems that can be discretized is the Branch & Prune (BP) algorithm. It recursively constructs the discretized search domain (a tree) and verifies the feasibility of the computed atomic positions. When all available distances are exact, all candidate positions for a given atom can be enumerated. This is however not possible in presence of interval distances, because a continuous subset of positions can actually be computed for the corresponding atoms. The focus of the work in [22] is on a new scheme for an adaptive generation of a discrete subset of candidate positions from this continuous subset. The generated candidate positions do not only satisfy the distances employed in the discretization process, but also additional distances that might be available (the so-called pruning distances).

Since the BP algorithm can loose in performance when dealing with large molecules containing several interval distances, we worked this year on a variation of the algorithm named BetaMDGP [29]. This is a work in collaboration with Korean researchers. The BetaMDGP algorithm is based on the concept of beta-complex, which is a geometric construct extracted from the quasi-triangulation derived from the Voronoi diagram of atoms.

From the theoretical side, we worked on two main directions. First, we proved that, in discretizable distance geometry problems where all available distances are exact, the total number of solutions is always a power of two. This is related to the fact that the discrete search space contains several symmetries [18]. Secondly, we tried to summarize in [37] the current issues for efficiently solving real-life instances of the distance geometry.

Finally, the work we performed during the last years, including another important results from other colleagues currently working on this topic, was summarized in an extensive survey on the discretization of the distance geometry [17].

Distance measure between Protein structure

We propose here a new distance measure for comparing two protein structures based on their contact map representations (CMO). This novel measure (max-CMO metric), satisfies all properties of a metric on the space of protein representations. Having a metric in that space allows to avoid pairwise comparisons on the entire database and thus to significantly accelerate exploring the protein space compared to non metric spaces. We show on a gold-standard classification benchmark sets that our exact k-nearest neighbor scheme classifies up to 95% and 99% of queries correctly. Our k-NN classification thus provides a promising approach for the automatic classification of protein structures based on contact map overlap. [26], [30]

Local similarity of protein structure

Finding similarities between protein structures is a main goal in molecular biology. Most of the existing tools preserve order and only find single alignments even when multiple similar regions exist. We propose a new seed-based approach that discovers multiple pairs of similar regions. Its computational complexity is polynomial and it comes with a quality guarantee that the returned alignments have both Root Mean Squared Deviations (coordinate-based as well as internal-distances based) lower than a given threshold, if such exists. We do not require the alignments to be order preserving, which makes our algorithm suitable for detecting similar domains when comparing multi-domain proteins. And because the search space for non-sequential alignments is much larger than for sequential ones, the computational burden is addressed by using both a coarse-grain level parallelism and a fine-grain level parallelism. [33]

IBIS Project-Team

5. New Results

5.1. Highlights of the Year

A paper based on the PhD thesis of Diana Stefan was accepted for *PLoS Computational Biology* this year [7].

5.2. Analysis of gene regulatory networks by means of piecewise-linear (PL) models

GENETIC NETWORK ANALYZER (GNA) is a tool for the qualitative modeling and simulation of the dynamics of gene regulatory networks by means of PLDE models, as described in Section 4.1 . GNA has been integrated with the other bioinformatics tools distributed by Genostar (http://www.genostar.com/). Version 8.7 of GNA was released by IBIS and Genostar this year. This version is an update of version 8.0, deposited at the Agence pour la Protection des Programmes (APP). Some bugs have been corrected in the new version and the program has been adapted to the latest versions of Java and the software platform of Genostar. Version 8.7 supports the SBML standard and is also capable of exporting its models to the newly-developed standard for qualitative models, SBML Qual. This standard has been elaborated by the community of developers of logical and related modeling tools (CoLoMoTo), in which the GNA developers participate.

The predictions obtained with the help of GNA are purely qualitative, describing the dynamics of the network by means of a state transition graph. While a qualitative analysis is appropriate for certain problems, the absence of precise quantitative predictions may not be desirable in others, such as the analysis of a limit cycle or the design of a controller for a synthetic network. The quantitative study of PLDE models of gene regulatory networks is hindered by the fact that the step functions describing the logic of regulatory interactions lead to discontinuities in the right-hand side of the PLDE models (Section 3.1). This has motivated extensions of the PLDE models based on differential inclusions and Filippov solutions. As of now, no numerical simulation tool for the simulation of these Filippov extensions is available.

In collaboration with the BIPOP project-team, we have shown how tools developed for the simulation of nonsmooth mechanical, electrical and control systems can be adapted for this purpose, in a paper published in *Physica D* [1] and presented at the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014) [12]. We have presented a method for the numerical analysis of one proposed extension, called Aizerman–Pyatnitskii (AP)-extension, by reformulating the PLDE models as mixed complementarity systems (MCSs). This allows the application of powerful methods developed for this class of nonsmooth dynamical systems, in particular those implemented in the SICONOS platform developed by BIPOP. We have also shown that under a set of reasonable biological assumptions, putting constraints on the right-hand side of the PLDE models, AP-extensions and classical Filippov extensions are equivalent. This means that the proposed numerical method is valid for a range of different solution concepts. We have illustrated the practical interest of our approach through the numerical analysis of three well-known networks developed in the field of synthetic biology.

5.3. Inference of bacterial regulatory networks from reporter gene data

The use of fluorescent and luminescent reporter genes allows real-time monitoring of gene expression, both at the level of individual cells and cell populations (Section 3.2). In order to fully exploit this technology, we need methods to rapidly construct reporter genes, both on plasmids and on the chromosome, mathematical models to infer biologically relevant quantities from the primary data, and computer tools to achieve this in an efficient and user-friendly manner. For instance, in a typical microplate experiment, 96 cultures are followed in parallel, over several hours, resulting in 10,000-100,000 measurements of absorbance and fluorescence and luminescence intensities. Over the past few years, we put into place an experimental platform and data analysis software, notably the WELLREADER program (Section 4.2), to allow biologists to make the most out of the information contained in reporter gene expression data.

Valentin Zulkower, in the framework of his PhD thesis, has developed novel methods for the analysis of reporter gene data, based on the use of regularized linear inversion. This allows a range of estimation problems in the analysis of reporter gene data, notably the inference of growth rate, promoter activity, and protein concentration profiles, to be solved in a mathematically sound and practical manner. We have evaluated the validity of the approach using *in-silico* simulation studies, and observed that the methods are more robust and less biased than indirect approaches usually encountered in the experimental literature based on smoothing and subsequent processing of the primary data, like in WELLREADER. We have applied the methods to the analysis of fluorescent reporter gene data acquired in kinetic experiments with *Escherichia coli*. The methods were shown capable of reliably reconstructing time-course profiles of growth rate, promoter activity, and protein concentration from weak and noisy signals at low population volumes. Moreover, they captured critical features of those profiles, notably rapid changes in gene expression during growth transitions. The linear inversion methods have been implemented in the Python package WELLFARE, and integrated by Michel Page in the web application WELLINVERTER (Section 4.2). This work was submitted for publication early 2015.

The above tools have been used in a series of studies directed at the experimental mapping of gene regulatory networks in *E. coli*. A first example is a study, led by Stéphan Lacour in collaboration with Akira Ishihama and Hiroshi Ogasawara in Japan, on the lifestyle adaptation of *E. coli*. The study concerns the switch between swimming motility and biofilm formation in response to changes in environmental growth conditions. The stationary phase sigma factor RpoS is an important regulator of this switch since it stimulates adhesion and represses flagellar biosynthesis. By measuring the dynamics of gene expression, we show that RpoS inhibits the transcription of the flagellar sigma factor, FliA, in exponential growth phase. RpoS also partially controls the expression of CsgD and CpxR, two transcription factors important for bacterial adhesion. We have demonstrated that these two regulators repress the transcription of *fliA*, *flgM* and *tar*, and that this regulation is dependent on the growth medium. CsgD binds to the flgM and fliA promoters around their -10 promoter element, strongly suggesting direct repression. The results show that CsgD and CpxR also affect the expression of other known modulators of cell motility. An updated structure of the regulatory network controlling the choice between adhesion and motility was proposed in the paper based on this work, published in the *Journal of Bacteriology* [2]. Stéphan Lacour also reviewed this and other work on RpoS in a publication in *Environmental Microbiology Reports* [4].

A second example derives from the PhD thesis of Diana Stefan. Although from a biological point of view the motility network of *E. coli* is also central in this work, its main thrust lies in clarifying and solving methodological issues in the automated inference of quantitative models of gene regulatory networks from time-series gene expression data, also called reverse engineering in the bioinformatics literature. The application of existing reverse engineering methods is commonly based on implicit assumptions on the biological processes under study. First, the measurements of mRNA abundance obtained in transcriptomics experiments are taken to be representative of protein concentrations. Second, the observed changes in gene expression are assumed to be solely due to transcription factors and other specific regulators, while changes in the activity of the gene expression machinery and other global physiological effects are neglected. While convenient in practice, these assumptions are often not valid and bias the reverse engineering process. In her PhD thesis, Diana Stefan systematically investigated, using a combination of models and experiments, the importance of this bias and possible corrections. She measured with the help of fluorescent reporter genes the activity of genes involved in the FliA-FlgM module of the *E. coli* motility network. From these data, protein concentrations and global physiological effects were estimated by means of kinetic models of gene expression. The results indicate that correcting for the bias of commonly-made assumptions improves the quality of the models inferred from the data. Moreover, it was shown by simulation that these improvements are expected to be even stronger for systems in which protein concentrations have longer half-lives and the activity of the gene expression machinery varies more strongly across conditions than in the FliA-FlgM module. The approach proposed in this study is broadly applicable when using time-series transcriptome data to learn about the structure and dynamics of regulatory networks. The paper describing the work was published in *PLoS Computational Biology* [7].

5.4. Models of carbon metabolism in bacteria

All free-living bacteria have to adapt to a changing environment. Specific regulatory systems respond to particular stresses, but the most common decision bacteria have to make is the choice between alternative carbon sources, each sustaining a specific, maximal growth rate. Many bacteria have evolved a strategy that consists in utilizing carbon sources sequentially, in general favouring carbon sources that sustain a higher growth rate. As long as a preferred carbon source is present in sufficient amounts, the synthesis of enzymes necessary for the uptake and metabolism of less favourable carbon sources is repressed. This phenomenon is called Carbon Catabolite Repression (CCR) and the most salient manifestation of this regulatory choice is diauxic growth, a phenomenon discovered by Jacques Monod more than 70 years ago. Although this system is one of the paradigms of the regulation of gene expression in bacteria, the underlying mechanisms remain controversial. Carbon catabolite repression involves the coordination of different subsystems of the cell - responsible for the uptake of carbon sources, their breakdown for the production of energy and precursors, and the conversion of the latter to biomass.

The complexity of this integrated system, with regulatory mechanisms cutting across metabolism, gene expression, signaling and subject to global physical and physiological constraints, has motivated important modeling efforts over the past four decades, especially in the enterobacterium *Escherichia coli*. Different hypotheses concerning the dynamic functioning of the system have been explored by a variety of modeling approaches. In an article in *Trends in Microbiology* [3], which was initiated during the sabbatical of Andreas Kremling in Grenoble in 2013, we have reviewed these studies and summarized their contributions to the quantitative understanding of carbon catabolite repression, focusing on diauxic growth in E. coli. Moreover, we have proposed a highly simplified representation of diauxic growth that makes it possible to bring out the salient features of the models proposed in the literature and confront and compare the explanations they provide.

A bottleneck in the development of dynamic and quantitatively predictive models of bacterial metabolism, explicitly accounting for the different regulatory mechanisms on the molecular level, is information on the kinetic parameters describing the enzymatic reactions and other molecular interactions. One particularly important piece of information is knowledge of enzyme concentrations. Recent technological advances in quantitative proteomics have made mass spectrometry-based quantitative assays an interesting alternative to more traditional immuno- affinity based approaches for quantifying enzyme concentrations. In particular, these advances have improved specificity and multiplexing capabilities. In a study carried out at CEA Grenoble, a quantification workflow to analyze enzymes involved in central metabolism in *E. coli* was developed. This workflow combined full-length isotopically labeled standards with selected reaction monitoring analysis. The workflow was used to accurately quantify 22 enzymes involved in *E. coli* central metabolism in a wild-type reference strain and two derived strains, optimized for higher NADPH production. Delphine Ropers and Hidde de Jong participated in the analysis of these data. In combination with measurements of metabolic fluxes, we showed that proteomics data can be used to assess different levels of regulation, in particular enzyme abundance and catalytic rate. This is key to the development of predictive kinetic models, but also provides information that can be used for strain design in biotechnology. An article based on this work was published in *Molecular and Cellular Proteomics* [8].

Other ongoing work on the analysis of bacterial metabolism is carried out by Delphine Ropers in collaboration with Inra/INSA in Toulouse, in the framework of the PhD thesis of Manon Morin, supported by a Contrat Jeune Scientifique Inra-Inria. In their respective PhD theses, Stéphane Pinhal and Valentin Zulkower also study specific aspects of carbon metabolism, using both models and experimental data.

5.5. Stochastic modeling and identification of gene regulatory networks in bacteria

At the single-cell level, the processes that govern single-cell dynamics in general and gene expression in particular are better described by stochastic models. Modern techniques for the real-time monitoring of gene expression in single cells enable one to apply stochastic modelling to study the origins and consequences of random noise in response to various environmental stresses, and the emergence of phenotypic variability. The potential impact of single-cell stochastic analysis and modelling ranges from a better comprehension of the biochemical regulatory mechanisms underlying cellular phenotypes to the development of new strategies for the (computer assisted or genetically synthesized) control of cell populations and even of single cells.

Work in IBIS on the probabilistic gene expression and interaction dynamics at the level of individual cells is addressed in terms of identification of intrinsic noise models from population snapshot data, on the one hand, and the inference of models focusing on cellular variability within isogenic populations from individual cell fluorescence microscopy gene expression profiles, on the other hand. Along with modelling and inference comes analysis of the inferred models in various respects, notably in terms of single-cell state estimation and control. Other problems related with single-cell modelling and extracellular variability are considered in high-eukariotic cells through external collaborations.

In the context of yeast cell response to osmotic shocks, in collaboration with the CONTRAINTES project-team, and colleagues from Université Paris Descartes and University of Pavia (Italy), Eugenio Cinquemani has investigated the use of mixed effects-modelling and identification techniques to characterize individual cell dynamics in isogenic cell populations. Mixed-effects models are hierarchical models where parametric response profiles of individuals is subject to inter-individual parameter variability following a common population distribution. Starting from identification approaches in pharmacokinetics, we have developed and applied inference methods to the context of microfluidics data, with focus on the budding yeast response to osmotic shocks. First results presented at conference in 2013 have been taken further, both in terms of mathematical analysis of the models developed and in terms of biological interpretation. Model identification and validation were performed together with Andres Gonzales, PhD student at the University of Pavia, who has visited IBIS for six months in 2014. A journal publication is currently being prepared for publication.

In a second line of work, starting from the models inferred in the above collaboration, the problem of real-time state estimation and control of single yeast cells has been considered. Together with the BIOCORE project-team, we have put in place algorithms for state estimation in presence of hybrid random switching and continuous dynamics, and integrated them with a feedback control approach developed by collaborators at TU Delft (the Netherlands). The whole monitoring, estimation and control chain has been deployed and applied *in silico* to the stochastic control of osmosensitive genes in single yeast cells. Methods and results have been presented at the 12th international conference on Computational Methods for Systems Biology (CMSB 2014), whose proceedings have been published as a volume of the LNCS series [14]. It is shown in particular that stochastic model-based estimation and control outperforms existing methods of single-cell control based on deterministic approximations.

Additional work on identification and estimation of hidden states for intrinsic noise models of gene expression/regulation in single bacterial cells, started with reference to arabinose uptake dynamics but also applicable to other regulatory networks in *E. coli*, is being developed. In parallel, collaboration of Eugenio Cinquemani with Marianna Rapsomaniki, PhD student affiliated with the University of Patras (Greece) and ETH Zürich (Switzerland), has been devoted to the analysis of data from Fluorescence Recovery After Photobleaching (FRAP) experiments and the inference of kinetic parameters of protein dynamics in single high-eukariotic cells. As an alternative to current approximate analytical methods, we have explored inference methods based on simulation of biological processes in realistic environments at a particle level. We introduced and demonstrated a new method for the inference of kinetic parameters of protein dynamics, where a limited number of *in-silico* FRAP experiments is used to construct a mapping from FRAP recovery curves to the parameters sought. Parameter estimates from experimental data are then computed by applying the mapping to the observed recovery curves, at virtually no additional price for any number of experiments, along with the application of a bootstrap procedure for determining identifiability of the parameters and confidence intervals for their estimates. After validation on synthetic data, the method was successfully applied to the analysis of the nuclear proteins Cdt1, PCNA and GFPnls in mammalian cells, also shedding light on cell-to-cell variability of the protein kinetics. Method and results have recently been published in *Bioinformatics* [6].

5.6. Growth control in bacteria and biotechnological applications

A bacterial cell adapts its growth rate and the level of gene expression required to sustain growth to the environment, notably to the availability of nutrients providing the molecular building blocks and the energy required for growth. This adaptive response involves the global physiological state of the cell, in particular the activity of the gene expression machinery, and DNA-binding transcription factors and other specific regulators. While many studies have focused on networks of transcription factors, the analysis of the relative contributions of both transcription factors and global effects of the physiological state has received relatively little attention thus far. There is a huge literature on the molecular mechanisms coupling the activity of the gene expression machinery to changes in the nutritional quality of the environment, but a quantitative and dynamic picture of this very complicated regulatory system is still missing. Delphine Ropers and Edith Grac as well as Nils Giordano are developing models to achieve this, from bottom-up and top-down perspectives, respectively.

The quantitative models adopting the bottom-up pespective describe the molecular mechanisms controlling the activity of the gene expression machinery. The calibration and analysis of these models is made difficult by their complexity, the nonidentifiability of many parameter values, and the heterogeneity of experimental data sources. To overcome these difficulties, Delphine Ropers and Edith Grac are developing model ensembles with the same structure but different parameter values that are consistent with the experimental data. In collaboration with Jean-Luc Gouzé and Ismail Belgacem from the BIOCORE project-team at Inria Sophia-Antipolis-Méditerranée, they have analysed the dynamical behavior of a central module of these models, which controls the cellular concentration of the RNA polymerase, the key player of the transcriptional machinery. By means of model reduction approaches and monotone system theory, they have analyzed the equilibria of the system and their stability, which they could relate to biological observations on *E. coli*. This work has been published in the proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014) [9] and the 53rd IEEE Conference on Decision and Control (CDC 2014) [10]. A journal article is in preparation.

In the context of the PhD thesis of former IBIS member Jérôme Izard, we have studied the relation between the gene expression machinery, the global physiology of the cell, and the growth rate from a different perspective. Our aim was to change the mechanisms regulating the activity of the gene expression machinery in such a way so as to be able to externally control the growth rate of the cell. More precisely, we have engineered an *E. coli* strain in which the transcription of an essential component of the global gene expression machinery is under the tight control of an inducible promoter. By adjusting the inducer concentration in the medium we can adjust the activity of the gene expression machinery and thereby reversibly switch the growth rate of the bacterium between zero and the maximal growth rate. Our modified *E. coli* strain, described in a paper prepared for submission, opens new perspectives for studying the mechanisms of growth control as well as for developing biotechnological applications, the subject of the post-doctoral fellowship of Cindy Gomez Balderas-Barillot. We have submitted a patent proposing such applications, which underlies the technology transfer activities undertaken in the recently-started Reset project (Section 7.1).

LIFEWARE Team

6. New Results

6.1. Highlight: Xavier Duportet laureate of the AEF docteurs-entrepreneurs prize

Xavier Duportet has been awarded the AEF prize at the docteurs-entrepreneurs competition. His thesis, made jointly within Lifeware and the Weiss lab at MIT, was entitled "Developing new tools and platforms for mammalian synthetic biology: from the assembly and chromosomal integration of large genetic circuits to the engineering of artificial intercellular communication systems". He published his research in *Nucleic Acids Research* and *Nature Biotechnology* [7], [5]. In particular, he demonstrated the assembly and chromosomal integration in mammalian cells of the largest gene circuit integrated to date. Subsequently, he co-founded the startup company PhageX. He was also a laureate of the Concours National de Création d'Entreprises Innovantes and the Concours Mondial d'Innovation (personalized medicine track).

He is the president of the Hello Tomorrow challenge and vice-president of the Osons La France initiative. He has notably been featured in articles published in Le Monde, L'Obs, and L'Opinion. He has been an invited speaker at the prestigious 4th Congreso del Futuro in Santiago (Chili). [7]

6.2. Highlight: François Fages laureate of the French Academy of Sciences

François Fages was very honoured to receive the Michel Monpetit prize 2014 of the French Academy of Sciences for his contributions to fundamental computer science (unification theory and constraint logic programming) and computational systems biology (modeling of biochemical networks and design and supervision of the implementation of the BIOCHAM software).

6.3. Highlight: Pauline Traynard Best Student Paper Prize at CMSB 2014, for Trace Simplifications preserving Temporal Logic Formulae with Case Study in a Coupled Model of the Cell Cycle and the Circadian Clock

Participants: François Fages, Sylvain Soliman, Pauline Traynard.

Pauline Traynard was very pleased to receive the Best Student Paper Prize of the twelth International Conference on Computational Methods for Systems Biology, 17-19 November 2014, Univ. of Manchester, UK, for a communication on trace simplifications preserving temporal logic properties [19].

Calibrating dynamical models on experimental data time series is a central task in computational systems biology. When numerical values for model parameters can be found to fit the data, the model can be used to make predictions, whereas the absence of any good fit may suggest to revisit the structure of the model and gain new insights in the biology of the system. Temporal logic provides a formal framework to deal with imprecise data and specify a wide variety of dynamical behaviors. It can be used to extract information from numerical traces coming from either experimental data or model simulations, and to specify the expected behaviors for model calibration. The computation time of the different methods depends on the number of points in the trace so the question of trace simplification is important to improve their performance. In [19] we study this problem and provide a series of trace simplifications which are correct to perform for some common temporal logic formulae. We give some general soundness theorems, and apply this approach to period and phase constraints on the circadian clock and the cell cycle. In this application, temporal logic patterns are used to compute the relevant characteristics of the experimental traces, and to measure the adequacy of the model to its specification on simulation traces. Speed-ups by several orders of magnitude are obtained by trace simplification even when produced by smart numerical integration methods.

6.4. Highlight: Modeling Dynamics of Cell-to-Cell Variability in TRAIL-induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance

Participants: Grégory Batt, François Bertaux, Szymon Stoma.

Isogenic cells sensing identical external signals can take markedly different decisions. Such decisions often correlate with pre-existing cell-to-cell differences in protein levels. When not neglected in signal transduction models, these differences are accounted for in a static manner, by assuming randomly distributed initial protein levels. However, this approach ignores the *a priori* non-trivial interplay between signal transduction and the source of this cell-to-cell variability: temporal fluctuations of protein levels in individual cells, driven by noisy synthesis and degradation. Thus, modeling protein fluctuations, rather than their consequences on the initial population heterogeneity, would set the quantitative analysis of signal transduction on firmer grounds. Adopting this dynamical view on cell-to-cell differences amounts to recast extrinsic variability into intrinsic noise. In collaboration with Dirk Drasdo (EPI Mmaba), we proposed a generic approach to merge, in a systematic and principled manner, signal transduction models with stochastic protein turnover models. When applied to an established kinetic model of TRAIL-induced apoptosis, our approach markedly increased model prediction capabilities [4]. We obtained a mechanistic explanation of yet-unexplained observations on fractional killing and non-trivial robust predictions of the temporal evolution of cell resistance to TRAIL in HeLa cells. Our results provide an alternative explanation to survival via induction of survival pathways since no TRAIL-induced regulations are needed and suggest that short-lived anti-apoptotic protein Mcl1 exhibit large and rare fluctuations. More generally, our results highlight the importance of accounting for stochastic protein turnover to quantitatively understand signal transduction over extended durations, and imply that fluctuations of short-lived proteins deserve particular attention. [4]

6.5. Towards Real-time Control of Gene Expression at the Single Cell Level: A Stochastic Control Approach

Participants: Grégory Batt, Pascal Hersen.

Recent works have demonstrated the experimental feasibility of real-time gene expression control based on deterministic controllers. By taking control of the level of intracellular proteins, one can probe single-cell dynamics with unprecedented flexibility. However, single-cell dynamics are stochastic in nature, and a control framework explicitly accounting for this variability is presently lacking. In [21], we devised a stochastic control framework, based on Model Predictive Control, which fills this gap.

Based on a stochastic modelling of the gene response dynamics, our approach combined a full state-feedback receding-horizon controller with a real-time estimation method that compensated for unobserved state variables. Using previously developed models of osmostress-inducible gene expression in yeast, we showed *in silico* that our stochastic control approach outperformed deterministic control design in the regulation of single cells. This contribution lead to envision the application of the proposed framework to wet lab experiments in yeast.

This work was done in collaboration with Alfonso Carta (EPI BIOCORE), Eugenio Cinquemani (EPI IBIS), Lakshmeesh Maruthi and Ilya Tkachev (TU Delft), and Alessandro Abate (Oxford U).

6.6. A Platform for Rapid Prototyping of Synthetic Gene Networks in Mammalian Cells

Participants: Grégory Batt, Xavier Duportet, Pascal Hersen.

Mammalian synthetic biology may provide novel therapeutic strategies, help decipher new paths for drug discovery and facilitate synthesis of valuable molecules. Yet, our capacity to genetically program cells is currently hampered by the lack of efficient approaches to streamline the design, construction and screening of synthetic gene networks. To address this problem, we developed a framework for modular and combinatorial assembly of functional (multi)gene expression vectors and showed their efficient and specific targeted integration into a well-defined chromosomal context in mammalian cells.

In [7], in collaboration with the Weiss lab and the MSC lab, we demonstrated the potential of this framework by assembling and integrating different functional mammalian regulatory networks including the largest gene circuit built and chromosomally integrated to date (6 transcription units, 27kb), encoding an inducible memory device. Using a library of 18 different circuits as a proof of concept, we also demonstrated that our method enabled one-pot/single-flask chromosomal integration and screening of circuit libraries. This rapid and powerful prototyping platform is well suited for comparative studies of genetic regulatory elements, genes and multi-gene circuits as well as facile development of libraries of isogenic engineered cell lines.

6.7. Reconfigurable Circuitry in Biochemical Systems

Participants: Hui-Ju Chiang, François Fages, Sylvain Soliman.

Realizing complex systems within a biochemical environment is a common pursuit in synthetic biology. Such systems achieve certain computation through properly designed biochemical reactions. Despite fruitful progress being made, most existing reaction designs have fixed target functionality. Their lack of reconfigurability can be disadvantageous, especially when a system has to adapt to a varying biochemical environment.

When control systems are of concern, linear control is one of the most widely applied control methods. Any linear control system can be realized with three elementary building blocks: integration, gain, and summation. Realizing linear control with biochemical reactions has been proposed in previous work, where reaction rates of the underlying reactions play a key role to achieve the desired building blocks. Essentially the reaction rates have to be matched exactly, and it imposes serious practicality restriction because in reality the reaction rates of available reactions are predetermined and can be limited. In [16] we devise a mechanism to make linear control systems configurable by adding auxiliary species as control knobs. The concentrations of the auxiliary species can be adjusted not only to compensate reaction rate mismatch, but also to reconfigure different control systems out of the same control architecture.

Furthermore, in [15] we propose an analog approach to economically construct a reconfigurable logic circuit similar to a silicon based field programmable gate array (FPGA). The effective “logic” and “interconnect” of the circuit can be dynamically reconfigured by controlling the concentrations of certain knob species. We study a potential biomedical application of our reconfigurable circuitry to disease diagnosis and therapy at a molecular level.

6.8. Inferring Reaction Systems from Ordinary Differential Equations

Participants: François Fages, Steven Gay, Sylvain Soliman.

In Mathematical Biology, many dynamical models of biochemical reaction systems are presented with Ordinary Differential Equations (ODE). Once kinetic parameter values are fixed, this simple mathematical formalism completely defines the dynamical behavior of a system of biochemical reactions and provides powerful tools for deterministic simulations, parameter sensitivity analysis, bifurcation analysis, etc. However, without requiring any information on the reaction kinetics and parameter values, various qualitative analyses can be performed using the structure of the reactions, provided the reactants, products and modifiers of each reaction are precisely defined. In order to apply these structural methods to parametric ODE models, we study a mathematical condition for expressing the consistency between the structure and the kinetics of a reaction, without restricting to Mass Action law kinetics. This condition, satisfied in particular by standard kinetic laws, entails a remarkable property of independence of the influence graph from the kinetics of the reactions. We derive from this study a heuristic algorithm which, given a system of ODEs as input, computes a system of reactions with the same ODE semantics, by inferring well-formed reactions whenever possible. We show how this strategy is capable of automatically curating the writing of ODE models in SBML, and present some statistics obtained on the model repository biomodels.net [8].

6.9. Model Reductions by Tropical Equilibration

Participants: François Fages, Sylvain Soliman.

Model reduction is a central topic in systems biology and dynamical systems theory, for reducing the complexity of detailed models, finding important parameters, and developing multi-scale models for instance. While singular perturbation theory is a standard mathematical tool to analyze the different time scales of a dynamical system and decompose the system accordingly, tropical methods provide a simple algebraic framework to perform these analyses systematically in polynomial systems. The crux of these methods is in the computation of tropical equilibrations. In [11] we show that constraint-based methods, using reified constraints for expressing the equilibration conditions, make it possible to numerically solve non-linear tropical equilibration problems, out of reach of standard computation methods. We illustrate this approach first with the detailed reduction of a simple biochemical mechanism, the Michaelis-Menten enzymatic reaction model, and second, with large-scale performance figures obtained on the http://biomodels.net webcite repository.

6.10. Model Reductions by Subgraph Epimorphisms

Participants: François Fages, Steven Gay, Thierry Martinez, Sylvain Soliman.

In [9] we follow another route based on a purely structural method and study the problem of deciding the existence of a subgraph epimorphism between two graphs. Our interest in this variant of graph matching problem stems from the study of model reductions in systems biology, where large systems of biochemical reactions can be naturally represented by bipartite digraphs of species and reactions. In this setting, model reduction can be formalized as the existence of a sequence of vertex deletion and merge operations that transforms a first reaction graph into a second graph. This problem is in turn equivalent to the existence of a subgraph (corresponding to delete operations) epimorphism (i.e. surjective homomorphism, corresponding to merge operations) from the first graph to the second. In this paper, we study theoretical properties of subgraph epimorphisms in general directed graphs. We first characterize subgraph epimorphisms (SEPI), subgraph isomorphisms (SISO) and graph epimorphisms (EPI) in terms of graph transformation operations. Then we study the graph distance measures induced by these transformations. We show that they define metrics on graphs and compare them. On the algorithmic side, we show that the SEPI existence problem is NP-complete by reduction of SAT, and present a constraint satisfaction algorithm that has been successfully used to solve practical SEPI problems on a large benchmark of reaction graphs from systems biology.

6.11. Temporal Logic Modeling of Dynamical Behaviors: First-Order Patterns and Solvers

Participants: François Fages, Pauline Traynard, Sylvain Soliman.

We have written a book chapter [20] to describe how quantitative temporal logic formulae can be used to formalize imprecise dynamical behaviors of biological systems, and how such a formal specification of experimental observations can be used to calibrate models to real data, in a more versatile way than with curve fitting algorithms, and with more efficient dedicated solvers than with generic temporal logic solvers.

Based on this article, we investigated the correctness of various trace simplification methods, as mentionned in the highlight section above [19].

6.12. Logical Modeling of the Mammalian Cell Cycle

Participants: François Fages, Pauline Traynard, Denis Thieffry.

The molecular networks controlling cell cycle progression in various organisms have been previously modelled, predominantly using differential equations. However, this approach meets various difficulties as one tries to include additional regulatory components and mechanisms. This led to the development of qualitative dynamical models based on Boolean or multilevel frameworks, which are easier to define, simulate, analyse and compose. In a poster presented at ECCB 2014, we revisit the Boolean model of Fauré et al. for the core network controlling G/S transition in mammalian cell cycle, taking into account recent advances in the characterisation of the underlying molecular networks to obtain a better qualitative consistency between model simulations and documented mutants features. In particular, we introduced Skp2, the substrate recruiting component of the SCFSkp2 complex, which targets cell cycle control elements, such as p27, and is repressed by the tumour suppressor protein Rb. Furthermore, to supersede the limitations inherent to the Boolean simplifications, we have considered the association of multilevel logical components with key cell cycle regulators, including the tumour suppressor protein Rb. Indeed, it is well established that differently phosphorylated forms of Rb result in different effects on other components of the network, which can be faithfully modelled using a multilevel rather than a Boolean variable. To evaluate the dynamical properties of the resulting models, we perform synchronous and asynchronous simulations using the software GINsim (http://www.ginsim.org), for both the wild-type case and documented perturbations (e.g. combinations of loss- or gain-of-function mutations). In addition, we have designed a series of temporal logic queries (expressed in the CTL language), which enable an efficient and automatic verification of key dynamical properties (existence of a cyclic attractor or of a stable state, conditions on the order of changes of component levels, etc.), using the popular symbolic model checker NuSMV. This strategy greatly facilitates the dynamical analysis of increasingly detailed and complex cell cycle models. Our goal is to obtain a core cell cycle model consistent with the most relevant experimental results on mammalian cells, which will then be used as a module in more comprehensive cellular models, including cross-talks with the circadian clock network and key signalling pathways, whose deregulation underlies the development of various cancers.

6.13. A Greedy Heuristic for Optimizing Metro Regenerative Energy Usage compared to CMA-ES and MILP

Participants: François Fages, David Fournier.

When the regenerative braking energy cannot be stored by the metro producing it, it has to be used instantaneously on the network, otherwise it is lost. In this case, the accelerating and braking trains need be synchronized to fully benefit from the regenerative energy, and a metro timetable is energetically optimized when all the regenerative braking is utilized to power other trains. This synchronization consists in lining up each braking train with an accelerating one in its neighbourhood. Doing so, the latter will benefit from the regenerative energy of the former. In [17], [3] a fast greedy heuristic is proposed to tackle the problem of minimizing the energy consumption of a metro timetable by modifying solely the dwell times in stations. This heuristic is compared to a state-of-the-art meta heuristic called the covariance matrix adaptation evolution strategy (CMA-ES) and shows similar results with much faster computation time. Finally, it is shown that a run of the algorithm on a full timetable may reduce its energy consumption by 5.1%.

MAGNOME Project-Team

6. New Results

6.1. Highlights of the Year

In collaboration with colleagues from the Institut du Vigne et du Vin (ISVV), Bordeaux and the Universidade Nova de Lisboa, Lisbon we used a population genomics approach to investigate the global phylogeography and domestication fingerprints of winemaking yeasts, using a collection of isolates obtained from fermented beverages and from natural environments on five continents. These results appeared in *Nature Communications* [11].

6.2. A Gondwanan imprint of S. uvarum diversity

Domestication of livestock and crops has been amply demonstrated through historical and archeological records, but domestication of microorganisms is much more difficult to establish. In a large-scale study [11] of the wine and cider yeast *Saccharomyces uvarum* conducted with colleagues from the Institut du Vigne et du Vin, Bordeaux and the Universidade Nova de Lisboa, Lisbon, we found the first indications of its domestication in the transition from its habitat in *Nothofagus* (southern beech) trees on the Gondwana mega-continent, to its present-day diversity in the Holarctic. The global phylogeography of these microorganisms was investigated through genome sequencing and comparison of 54 strains isolated on five continents, resulting in the identification of 10^5 high-quality SNPs and a remarkable pattern of introgressions ([11] figure 3 http://dx.doi.org/10.1038/ncomms5044).

The 54 genomes in this study were isolated, selected, and sequenced, and both assembled and aligned against reference genomes. Phylogenies were based on concatenated SNP alignment of selected chromosomes. The structure of the population was investigated using model-based Bayesian clustering.

In addition to the biological result, this study illustrates the ubiquity of an experimental approach based on large-scale sequencing of highly related genomes, in order to isolate tiny differences linked to a trait of interest. This is in contrast to the strategy that was current eight years ago, based on sequencing of a modest number of genomes spanning a much greater evolutionary range.

6.3. Improving inference of metabolic models

Participants: David James Sherman [correspondant], Pascal Durrens, Razanne Issa, Anna Zhukova.

The Pantograph approach uses reference model annotated by *gene associations*, and voting between complementary predictions of homology between reference genes and target genes, to decide whether a reaction that is present in the scaffold ought be be present in the target. A gene association implicitly represents expert knowledge about the role of genes in a compact way. If the gene association can be rewritten into a possibly satisfiable formula, then the corresponding reaction is instantiated in the target model.

Historically, gene associations have been used intuitively by experts during the model design and curation process, and are often inconsistent. We have formalized the construction of gene associations based on the semantics of different interpretations, showing how different boolean formulas should be constructed when the application is *i*) metabolic model inference, *ii*) flux-balance analysis, *iii*) hierarchical modeling, or *iv*) dynamic simulation (Razanne Issa, MS in prep.).

Second, we have refined our strategy for inferring metabolic models using abductive logic. We have shown that given a set of genes as observations in the target organism, and rules for rewriting gene associations while respecting integrity constraints for the model, then the reactions in the target model can be abduced as hypotheses that “explain” the presence of a maximial number of genes in the target genome. The advantage of this approach is that it can invent, through specialization, reactions that are not present *per se* in the reference model. Two classes of reactions can be invented: substrate-specific reactions inferred from expansion in gene families, and transport reactions needed to maintain model integrity for constitutive compartments.

6.4. Knowledge-based generalization of metabolic models

Participants: David James Sherman [correspondant], Pascal Durrens, Razanne Issa, Anna Zhukova.

Large metabolic networks are hard to understand and curate, because the large number of detailed reactions, which are needed for accurate modeling and simulation, obscure the high-level structure of the reaction network. We defined knowledge-based methods that factor similar reactions into “generic” reactions in order to visualize a whole pathway or compartment, while maintaining the underlying model so that the user can later “drill down” to the specific reactions if need be[15], [16] An implementation of this method is available as a Python library (see paragraph 5.3).

Figures 2 and 3 illustrate model generation for *Yarrowia lypolitica* fatty acid oxidation in the peroxisome. Molecular species are represented using SBGN notation: as circular nodes, and the reactions as square ones, connected by edges to their reactants and products. Ubiquitous species are of smaller size and colored gray. Non-ubiquitous species are divided into six equivalence classes, and coloured accordingly. The size of the model does not allow for readability of the species labels, thus we do not show them (figure 2).

The specific model is appropriate for simulation, because it contains all of the precise reactions. The generalized model is suited for a human, because it reveals the main properties of the model and masks distracting details. For example, the generalized model highlights the fact that there is a particularity concerning *C24:0-CoA (stearoyl-CoA)* (yellow): there exists a "shortcut" reaction (orange), producing it directly from another *fatty acyl-CoA* (yellow), avoiding the usual four-reaction beta-oxidation chain, used for other *fatty acyls-CoA*. This shortcut is not obvious in the specific model, because it is hidden among a plethora of similar-looking reactions.

We formally defined the generalization method in [15] and showed how to calculate it using a good approximation to an NP-complete set cover problem. The method was further validated in a collection of 1283 inferred models and revealed, on the one hand, a number of probable errors in the inferred models, and on the other hand, that there exist different families of generalization with a plausible link to different adaptive responses.

6.5. Characterization of STAND protein families

Participants: David James Sherman, Pascal Durrens, Witold Dyrka [correspondant].

In collaboration with Sven Saupe and Mathieu Paoletti from IBGC Bordeaux (ANR Mykimun), we worked on characterization of the STAND protein family in the fungal phylum. We established an *in silico* screen based on state-of-the-art bioinformatic tools, which – starting from experimentally studied sequences from *Podospora anserina* – allowed us to determine the first systematic picture of fungal STAND protein repertoire (ms. in preparation). Most notably, we found evidence of extensive modularity of domain associations, and signs of concerted evolution within the recognition domain [13]. Both results support the hypothesis that fungal STAND proteins, originally described in the context of vegetative incompatibility, are involved in a general fungal immune system. In addition, we investigated improved protein domain representations and elaborated a grammatical modelling method [23], which will be used to elucidate mechanisms of formation and operation of the STAND proteins.

NLR domains identified in this work have been incorporated into the upcoming release of Pfam [0].

To further explor the underlying mechanisms of repeat formation we implemented a stochastic string rewriting system that models the generation process of highly internally conserved repeats. The system is grounded in the biology of the process as it models transformation of repeats through the events of unequal crossing-over and mutation, which are believed to be main mechanisms that produce diversity in repeats. We confirmed that highly variable sites identified on the basis of entropy, are subject to selective pressure towards composition typical for binding sites, which is consistent with the suggested role of recognition epitopes.

---

[0] http://pfam.xfam.org

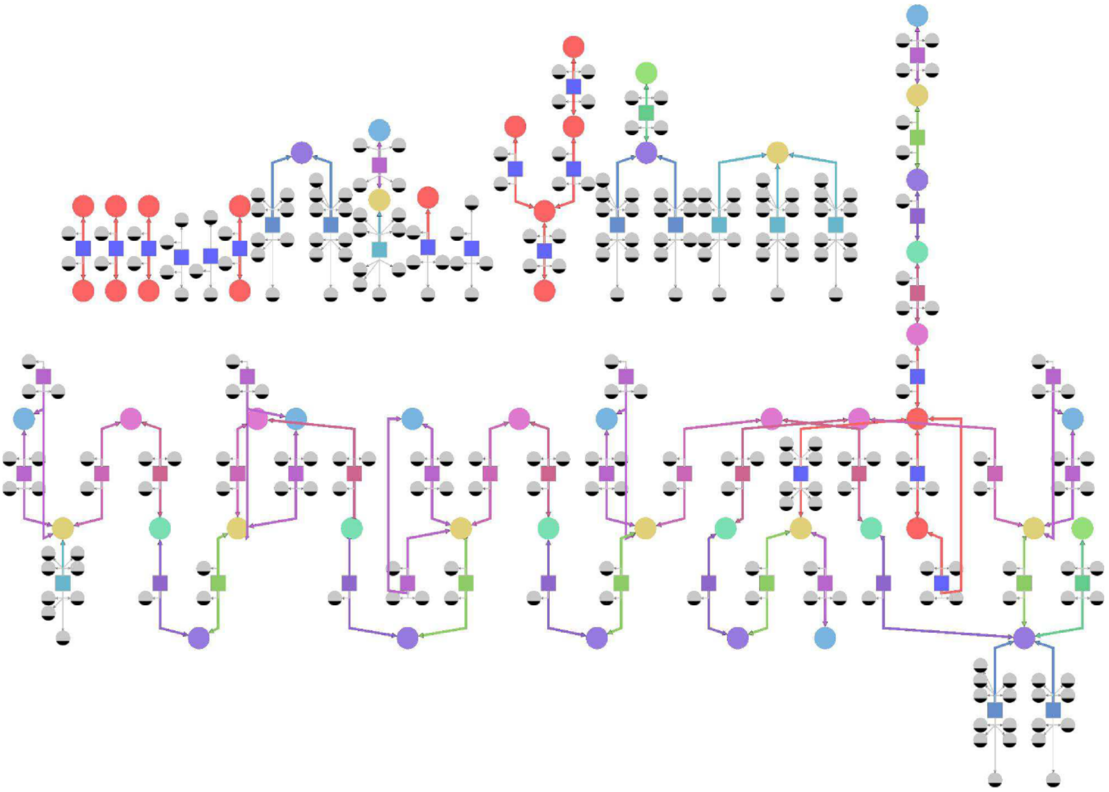


Figure 2. Yarrowia lypolitica fatty acid oxidation model before generalization. Reactions of the specific model are divided into fifteen equivalence classes, represented by different colours. Generally speaking, β-oxidation is a transformation of fatty acyl-CoA (yellow) into dehydroacyl-CoA (violet), then into hydroxyacy fatty acyl-CoA (dark green), 3-ketoacyl-CoA (magenta), and back to fatty acyl-CoA (with a shorter carbon chain); while the specific model describes the same process in more details, specifying those reactions for each of the fatty acyl-CoA species presented in the organisms' cell (e.g. decanoyl-CoA, dodecanoyl-CoA, etc.). This high-level, repetitive structure is obscured by the detail of the individual reactions.

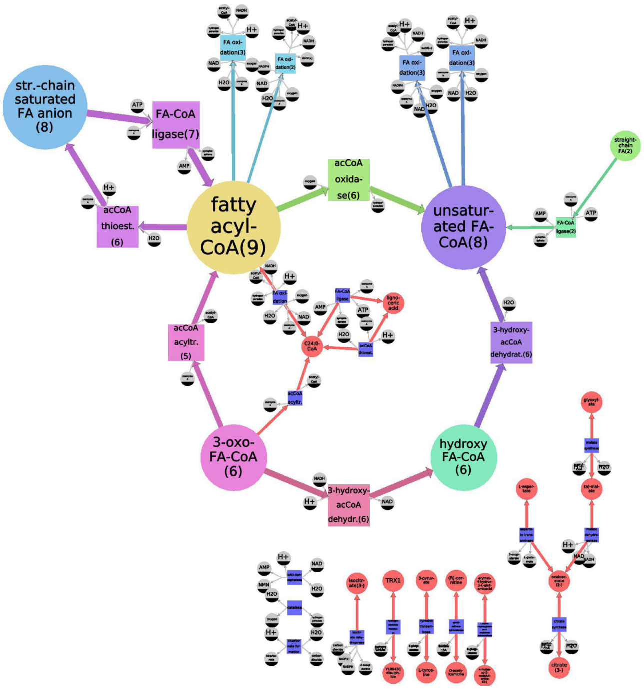


Figure 3. Generalization of the Yarrowia lypolitica fatty acid oxidation model, described as a transformation of fatty acyl-CoA (yellow) into dehydroacyl-CoA (violet), then into hydroxyacy fatty acyl-CoA (dark green), 3-ketoacyl-CoA (magenta), and back to fatty acyl-CoA with a shorter carbon chain. The generalization algorithm identifies equivalent molecular species using an ontology, and groups together reactions that operate on the same abstract species. It finds the greatest generalization the preserves stoichiometry. The generalized model represents quotient species and reactions. For example, the violet dehydroacyl-CoA node is a quotient of hexadec-2-enoyl-CoA, oleoyl-CoA, tetradecenoyl-CoA, trans-dec-2-enoyl-CoA, trans-dodec-2-enoyl-CoA, trans-hexacos-2-enoyl-CoA, trans-octadec-2-enoyl-CoA, and trans-tetradec-2-enoyl-CoA (colored violet in figure 2). In a similar manner, the light-green acyl-CoA oxidase quotient reaction, that converts fatty acyl-CoA (yellow) into dehydroacyl-CoA (violet), generalizes six corresponding light-green reactions of the initial model (figure 2).

MORPHEME Project-Team

5. New Results

5.1. Highlights of the Year

- Laure Blanc-Féraud was General Program chair of the conference IEEE ISBI 2014 in Beijing.

5.2. Sparse 3D reconstruction in fluorescence imaging

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Sébastien Schaub, Gilles Aubert.

Sparse reconstruction Super-resolution microscopy techniques allow to overstep the diffraction limit of conventional optics. Theses techniques are very promising since they give access to the visualisation of finer structures which is of fundamental importance in biology. In this work we deal with Multiple-Angle Total Internal Reflection Microscopy (MA-TIRFM) which allows reconstructing 3D sub-cellular structures of a single layer of $\sim 300\ nm$ behind the glass coverslip with a high axial resolution. The 3D volume reconstruction from a set of 2D measurements is an ill-posed inverse problem and requires some regularization. Our aim in this work is to propose a new reconstruction method for sparse structures that is robust to Poisson noise and background fluorescence. The sparse property of the solution can be seen as a regularizer using the ℓ_0-norm. Let us denote $\mathrm{f} \in \mathrm{R}^N$ the unknown fluorophore density, then the problem states as

$$\widehat{\mathrm{f}} = \arg \min_{\mathrm{f} \in \mathrm{R}^N} \left(\mathrm{J}_d(\mathrm{f}) + \lambda \|\mathrm{f}\|_0 \right) \tag{5}$$

where J_d is defined from the likelihood function of the observation given f, $\lambda > 0$ is a weight parameter and $\| \cdot \|_0$ denotes the ℓ_0-norm (which counts the number of nonzero components of f). In order to solve this combinatorial problem, we propose a new algorithm based on a smoothed ℓ_0-norm allowing minimizing the non-convex energy (1). Following [20], the idea is to approach the ℓ_0-norm by a suitable continuous function depending on a positive parameter and tending to the ℓ_0-norm when the parameter tends to zero. Then the algorithm solves a sequence of functionals which starts with a convex one (on a large convex set) and introduce progressively the non-convexity of the ℓ_0-norm (Graduated Non Convexity approach). Figure 1 shows the accuracy of the method on a simulated membrane.

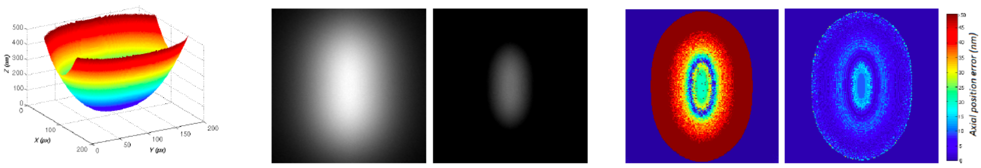


Figure 1. From left to right: Simulated membrane, Microscope acquisition (numerical simulations) with two different incident angles. The two images on the right represent position errors (nm) in the axial direction of the reconstructed membrane obtained with different algorithms: Richardson-Lucy algorithm without regularization (left) and our algorithm with $\lambda = 0.001$ (right).

Axial profile calibration In order to turn on real sample reconstructions we need to perform a calibration of the TIRF microscope. Its principle is based on an evanescent wave with an exponential theoretical decay. However this decay is generally not a pure exponential in practice and we need to have a good knowledge about it. Then based on a phantom specimen of known geometry (bead) we are working on a method to estimate experimentally/numerically this decay profile and calibrate all parameters of the system.

5.3. Penalty analysis for sparse solutions of underdeterminated linear systems of equations

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Gilles Aubert.

In many applications such as compression to reduce data storage, compressed sensing to recover a signal from fewer measurements, source separation, image decomposition and many others, one aims to compute a sparse solution of an underdetermined linear systems of equations. Thus finding such sparse solutions is currently an active research topic. This problem can be formulated as a least squares problem regularized with the ℓ_0-norm. We consider the penalized form

$$\hat{\mathrm{x}} = \arg \min_{\mathrm{x} \in \mathrm{R}^N} \left(\frac{1}{2} \|A\mathrm{x} - d\|^2 + \lambda \|\mathrm{x}\|_0 \right) \tag{6}$$

where $A \in \mathrm{R}^{M \times N}$, $d \in \mathrm{R}^M$ represents the data and $\lambda > 0$ is an hyperparameter characterizing the trade-off between data fidelity and sparsity.

It is well known that reaching a global solution of this $\ell_2 - \ell_0$ functional is a NP-hard combinatorial problem. Besides the non-convexity of this 'norm', its discontinuity at zero makes the minimization of the overall functional a hard task. In this work we focus on non-convex continuous penalties widely used to approximate the ℓ_0-norm which usually lead to better results than the classical ℓ_1 convex relaxation since they are more 'ℓ_0-like'. Based on some results in one dimension, we propose the Exact ℓ_0 penalty (El0). In one dimension and when the matrix A is orthogonal, replacing the ℓ_0-norm in (2) by this penalty gives the convex hull of the overall function. Then we have proved, for any matrix $A \in \mathrm{R}^{M \times N}$, that the global minimizers of the ℓ_2- El0 objective function are the same as for the $\ell_2 - \ell_0$ functional. We also demonstrate that all the local minimizers of this approximated functional are local minimizers for $\ell_2 - \ell_0$ while numerical experiments show that the reciprocal is in general false and that the objective function penalized with El0 admits less local minimizers than the $\ell_2 - \ell_0$ functional. Then, this work provides in some way an equivalence between the initial $\ell_2 - \ell_0$ problem and its approximation using the El0 penalty. One can address problem (2) by replacing the ℓ_0-norm with the El0 penalty which provides better properties for the objective function although the problem remains non-convex. Recently, some authors have proposed algorithms and proved their convergence to critical points of non-smooth non-convex functionals like ℓ_0-El0. Based on such algorithms, we propose a macro algorithm and prove its convergence to a (local) minimizer of the initial $\ell_2 - \ell_0$ functional.

5.4. Motion compensation in two-photon microscopy temporal series

Participants: Caroline Medioni, Grégoire Malandain, Florence Besse, Xavier Descombes.

Acquisitions of 3D image sequences over long period of time, in particular, have enabled neurobiologists to follow complex processes such as the development of neuronal populations or degenerative events occurring in pathological contexts, improving our understanding of the mechanisms involved in brain development and function. In most cases, live samples are moving/growing during long-term imaging. Therefore it is required to compensate for this global 3D motion before measuring the dynamics of the structure of interest. We have proposed a method to compute a coherent 3D motion over a whole temporal sequence of 3D volumes (Figure 2), which is able to capture subtle sub-voxelic displacements.

5.5. Axon Growth Imaging and Modeling

Participants: Agustina Razetti, Caroline Medioni, Florence Besse, Xavier Descombes.

The modeling part of this work has been made in collaboration with S. Komech, E. Pechersky and E. Zhizhina from IITP (Russian Academy of Science)

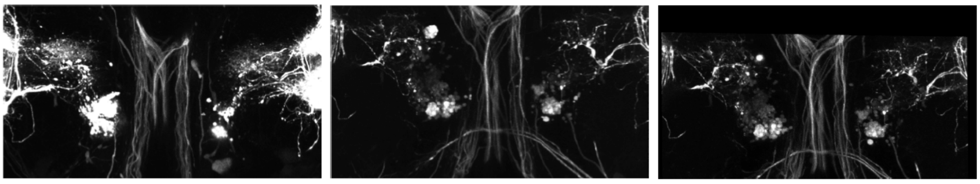


Figure 2. Left and middle: mip views of both the first and the last volumes of a temporal series. Right: mip view of the last volume after motion compensation.

In Drosophila brain, at metamorphosis, Mushroom Body gamma neurons undergo axonal remodeling characterized by a pruning of larval branches followed by regrowth and branching/arborization of adult processes. Axonal regrowth at this stage is essential to consolidate the adult brain and its success is determined by the trajectories followed by the axons and their branches. These trajectories depend on both extracellular guidance signals, and on a complex internal molecular machinery capable to read these signals and act in consequence. F. Besse's team at the IBV Institute has identified genes involved in this regrowth and branching processes [19]. A better understanding of the role of these genes will help to unravel the molecular mechanisms behind these fundamental processes, and lead to a better understanding of the neuronal morphology in both healthy and pathological conditions.

During this PhD project, mathematical and computational tools will be developed to characterize and compare the axonal regrowth and branching dynamics. Different populations of regrowing gamma axons will be considered (i.e. wild type and presenting mutations in relevant genes). The study will be based on both static 3D confocal images of axonal trees, and two-photon in vivo 4D image sequences showing either a single GFP positive regrowing axon or the entire population of regrowing gamma axons marked with GFP. In a first part we have begun to collect data. This includes three parts: i) fly stock maintenance, crossing and selection; ii) sample preparation consisting in pupal brain dissection, medium preparation and sample final assembly; iii) imaging: using different microscopy techniques, eg.confocal/two-photon, microscopy and light sheet microscopy, and acquisition of 4D image sequences. Thanks to an imaging technique developed by C. Medioni in F. Besse's laboratory, we have been able to monitor axonal regrowth and branching at early steps in individual neurons, and to follow them for about 15 hours using the two-photon microscope. These movies will be used in the future to develop the mathematical modeling of axonal regrowth/branching process (see figure 3). Our early works concerning modeling have consisted in investigating some models based on continuous time random walks and characterizing the main axon branch through topological entropy [8] [22].

5.6. Markov Chain for Axon Growth Modeling

Participants: Alejandro Mottini, Xavier Descombes, Florence Besse.

In this work we have defined a 2D discrete stochastic model for the simulation of axonal biogenesis [8]. The model is defined by a third order Markov Chain. The model considers two main processes: the growth process that models the elongation and shape of the neurites and the bifurcation process that models the generation of branches. The growth process depends, among other variables, on the external attraction field generated by a chemoattractant molecule secreted by the target area.

For the validation, we have fluorescently labeled single neurons within intact adult Drosophila fly brains, and have acquired 3D fluorescent confocal microscopy images of their axonal trees. Both normal neurons and neurons in which the function of the imp (mutant type 1) or profilin (mutant type 2) genes was inactivated were imaged. imp encodes a conserved RNA binding protein controlling subcellular mRNA transport and local protein synthesis, and is essential for axonal remodeling. profilin encodes a regulator of the actin cytoskeleton involved in axonal pathfinding. Mutations in these two conserved genes have been linked to neurological pathologies.

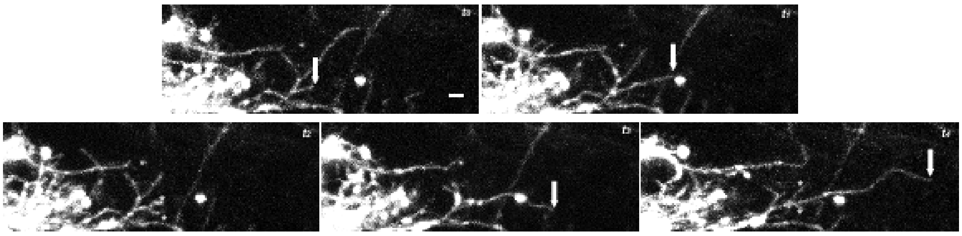


Figure 3. Images extracted from a movie recorded on a two-photon microscope (maximum intensity projection): single axon at the regrowing stage within Drosophila brain. Axons are marked with GFP. t0-t1: elongation step, t2: retraction and branching event, t3: elongation and t4: elongation and branching step. Arrows shows axonal tips and asterisks, the formation of branches. Scale bars for each image: 10μm.

Each image stack has a resolution of $0.093967 \times 0.093967 \times 0.814067 \mu m$ and two channels. The morphology of single axonal trees is visible in the first channel and was manually segmented by an experienced biologist. The morphology of the overall neuronal structure in which axons are developing is visible in the second channel. In total, 53 images (18 normal, 21 type 1 mutant and 14 type 2 mutant) were used. In order to study the attraction field of the populations, all stacks were registered against the first image of the normal population. This was performed using the second channel of each image.

We then have estimated the model parameters to generate two fields for each population, a scalar field that represents the axon flexibility and a vector field that represents the attraction field. Since we obtain some estimates on a sparse set of points in the x, y plane, we extrapolate the fields using a Gaussian Markov Random Field. By qualitatively analyzing the resulting images we have determined that there is no relevant difference on the attraction field between the three populations. We have observed that the field points towards the target area and that its norm is stronger at the starting point of the axons and weaker near the target area, which is consistent with biological expectations. The same procedure was used to analyze the difference on the scalar fields for each population (see Figure 4). In this case we can detect a significant difference between the populations.

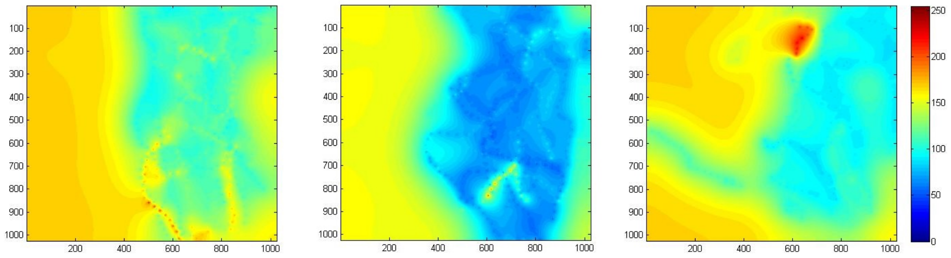


Figure 4. Markov Chain parameter scalar field for the normal (left), mutant type 1 (middle) and mutant type 2 (right) populations.

5.7. Cells detection using segmentation competition

Participants: Emmanuelle Poulain, Emmanuel Soubies, Sylvain Prigent, Xavier Descombes.

Image segmentation has been widely investigated in particular in the context of bioimaging for cells detection. In some cases, the background is clearly identifiable so that a binary mask of the objects can be computed using simple techniques such as thresholding. Therefore, isolated objects are easily recognizable while splitting clusters of objects, which are connected components in the binary mask, remains a challenging task. In fluorescent microscopy devices used for live imaging – *e.g.* confocal, biphoton, Selective Plane Illumination Microscope (SPIM) – an additional difficulty comes from the multiple degradations of the acquired images such as strong noise, spatially varying blur and light attenuation which makes the segmentation a hard task even for selecting a suitable threshold for the background. Since many years, researchers have developed several methods to perform such segmentation. An efficient approach consists in generating seeds that define regions using geometric information through a distance, as in the markers controlled watershed algorithm [21], or image gradient for the active contour approach. These approaches give accurate results providing the seeds are well chosen that is still an open issue. Bayesian approaches, such as marked point process, avoid this bottleneck by selecting randomly generated shapes through the minimization of an energy function. However, they are restricted to low dimensional parametric shapes, such as disks or ellipses, due to computational issues. Tuning the parameters of the segmentation algorithms mentioned above in order to obtain accurate results on the whole image can also be extremely tricky whereas it is much easier to obtain accurate results on different parts of the image using different sets of parameters. To overcome these limits we propose to combine both approaches by generating shapes from state of the art segmentation algorithms using random seeds and/or different sets of parameters. These shapes define a dictionary of candidates from which a competition process, using the Multiple Birth and Cut algorithm, extracts the most relevant shapes. We have validated this selection approach on synthetic data and on a multicellular tumor spheroide slice by comparing the obtained results with two different state of the art segmentation methods to build the dictionary of shapes and compare the performance of our competition approach with the ImageJ particle analyser (see figure 5).

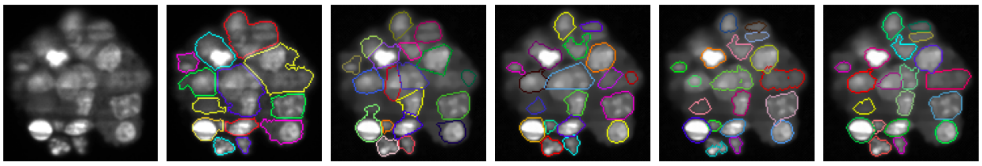


Figure 5. Spheroids of tumor cells stained with a fluorescent nuclear marker (top line left). Fiji Particle Analyzer result (top line right). Results of our approach with dictionary generated by RS-SKIZ with 800 repetitions and a threshold value fixed to 70 (middle line left) MTRS-SKIZ with 80 repetitions and 10 different thresholds (middle line right) MPRS-FM with 100 repetitions and 8 different values for the Fast Marching parameter used to stop the expansion (bottom line left), concatenation of MTRS-SKIZ and MPRS-FM dictionaries (bottom line right).

5.8. Graph cut and attractive interactions

Participants: Tarun Yellamraju, Emmanuel Soubies, Sylvain Prigent, Xavier Descombes.

Marked point processes have proved to be very efficient for segmenting a collection of objects from digital images. The multiple birth and death algorithm provides an optimization framework that allows reasonable time computation. This algorithm has to be embedded in a simluated annealing framework which involves parameters tuning (initial temperature and cooling scheme). This tedious task can be overcame considering a graph cut algorithm instead of the death step. The algorithm then consists in successively adding new random objects in the configuration and selecting the most relevant using the graph cut algorithm. In the graph construction a node is associated to each object. Unfortunately, the regularity condition imposed by the graph cut prevents to consider attractive interactions such as clustering or alignment constraints, which restricts the model to repulsive properties such as non overlap between objects. To overcome this restriction we have investigated new graph constructions by considering nodes defined by clusters of interacting objects. Different strategies have been compared to avoid being tracked in local minima defined by clusters while minimizing the number of required iterations. First results have been obtained on a seeds detection problem (see figure 6).

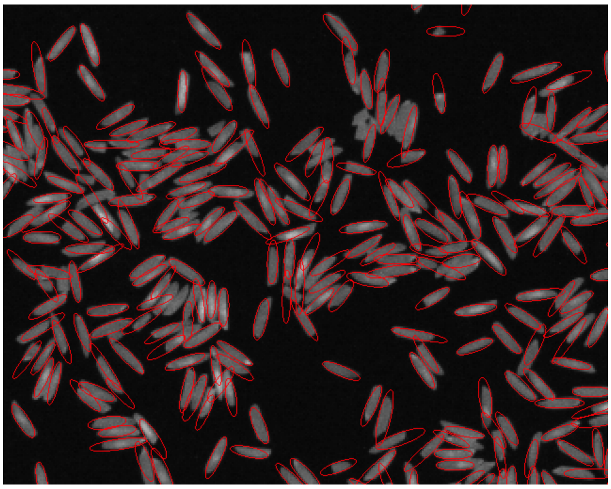


Figure 6. Rice seeds detection using a marked point processes and a birth and cut algorithm (Rice image has been given by Alpha MOS and LAAS-CNRS).

5.9. Cell-to-cell ascidian embryo registration

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Léo Guignard and Christophe Godin (Virtual Plants) and Patrick Lemaire (CRBM), within the Morphogenetics Inria Project Lab.

Recent microscopy techniques allow imaging temporal 3D stacks of developing organs or embryos with a cellular level of resolution and with a sufficient acquisition frequency to accurately track cell lineages. Imaging multiple organs or embryos in different experimental conditions may help to decipher the impact of genetic backgrounds and environmental inputs on the developmental program. For this, we need to precisely compare distinct individuals and to compute population statistics. The first step of this procedure is to develop methods to register individuals.

From a previous work of cell segmentation from microscopy images [6], we propose an approach to extract the Left-Right symmetry plane of embryos at early stages (Figure 7). Then we use the symmetry information to both register these embryos at a similar developmental stage and obtain a cell-to-cell mapping. We assessed the symmetry plane extraction on more than 100 images from 10 individuals between 32-cells and late-neurula development stage. The cell-to-cell registration was performed on 5 distinct individuals at 64-cells and 112-cells stage (Figure 8).

5.10. Quantitative comparison of micro-vasculatures

Participants: Manon Linder, Grégoire Malandain.

This work is made in collaboration with Cécile Duplaa and Thierry Couffinhal (INSERM).

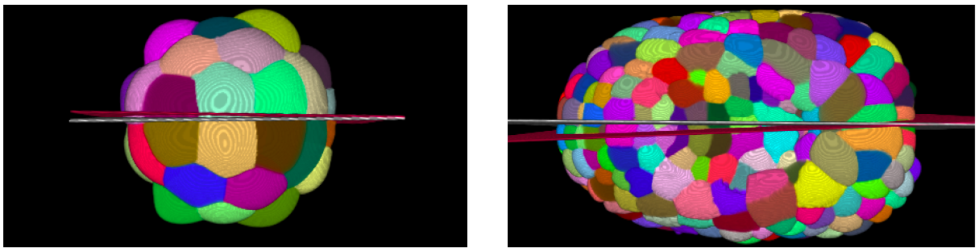


Figure 7. Left-Right symmetry plane initialization (red) and final estimation (white) on (left) a 32-cells stage embryo and (right) a neurula stage embryo.

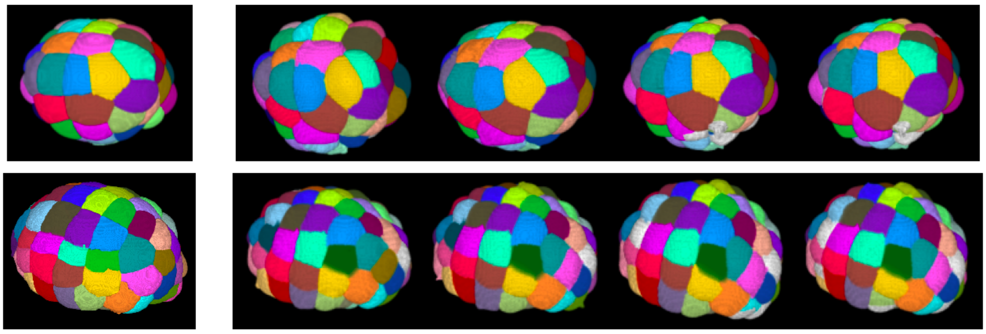


Figure 8. Cell-to-cell mapping between reference image (left) and test images (right) at 64-cells stage (first line) and at 112-cells stage (last line). The reference images are taken from the same individual, the test images are taken at different time-points of a second individual. On the test images, white cells are those that have not been matched to a reference cell.

Angiogenesis is a key component of ontogenesis, but also of tumor development or in some pathology repair (i.e. ischemia). Deciphering the underlying mechanisms of vessel formation is of importance. We aimed at identifying and characterizing the genetic components that are involved in this development. This requires to compare the effect of each gene with respect to the others, hence appeals for quantitative comparisons. We developed a methodology that first transforms a vascular image into a tree and second quantitatively analyze 3D vascular trees (see Figure 9) We conduct real experiments with images of the renal arterial network of different mutant mice, through the development of quantitative measurements that allow for group study.

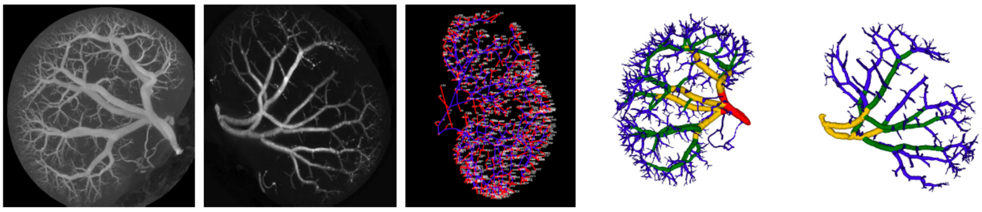


Figure 9. From left to right: two mip views of mouse kidneys acquired with a micro-CT (control and Fzd4 & Fzd6 KO mouse); a labeled tree built from an image; Diameter-defined Strahler classification of arterial trees.

5.11. Pre-clinical molecular dynamic image analysis: 99mTc- pertechnetate biodistribution model of murine stomach with micro-SPECT

Participants: Marine Breuilly, Grégoire Malandain.

This work is jointly conducted with Thierry Pourcher, Jacques Darcourt, Philippe Franken, Kaouthar Chatti, and Philippe Pognonec from the Transporter in Imagery and Oncologic Radiotherapy team (TIRO, CEA-CAL-UNSA).

This project investigates the potential retention of iodide in the stomach, for a better understanding of the iodide biodistribution in the body and more precisely of its potential antiseptic role. To that end, we study the uptake of the ^{99m}Tc-pertechnetate (an iodide analog) within the murine stomach observed thanks to a SPECT camera. Using the coupled SPECT and CT device dedicated to small animals, functional information targeted by a specific radiotracer (^{99m}Tc-pertechnetate) can be imaged dynamically.

The temporal evolution of the uptake is analysed thanks to a dedicated multi-compartment model. The addressed challenges consist in 1) estimating the time-activity curves for the different compartments, and 2) identifying the model parameters.

- ^{99m}-pertechnetate is an iodide analog regarding to the NIS gene. Thus iodide uptake kinetics can be studied through the study of ^{99m}Tc- pertechnetate biodistribution.
- Dynamic SPECT images exhibit a progressive accumulation of ^{99m}Tc-pertechnetate in the stomach wall and diffusion in the stomach cavity.
- The workflow that has been previously proposed in [18] was tested on a larger dataset of five subjects, yielding promising results: The computed model parameters are coherent, and the computed parameter values suggested that there is some iodide retention in the stomach wall.
- A comparison of the dedicated method for extraction of time activity curves with the ones extracted with Pixies software is on-going.
- A comparison of the dedicated method for solving the inverse problem of the compartmental analysis with methods developed by the Turku PET Centre is on-going.

5.12. Massal motility measures to automatically predict fertility scores

Participants: Ana Rita Lopes Simoes, Eric Debreuve.

This work has been done in the scope of the ANR project MOTIMO. We developed a method for automatic scoring of sperm samples in order to predict fertility for the farming industry. The method was applied to samples from rams and goats. A given sample is a video composed of a hundred frames (see Fig. 10).

We analyzed video samples acquired according to four modalities or protocols: drop, chamber, fluorescent beads with a 4x zoom, and fluorescent beads with 10x zoom. Two options have been considered. (1) An optical flow method has been applied to the videos in order to estimate the apparent motion of the seminal fluid (see Fig. 10). Some statistical features of interest (such as entropy) were extracted from the obtained motion fields in order to characterize the sperm massal motility. (2) The second option consisted of tracking the beads on the 4x or the 10x fluorescence videos (see Fig. 10). Some features of interest were also extracted from the resulting bead trajectories. Then using either of these feature sets (obtained with options (1) and (2)), a regression analysis (linear and kernel SVM) was conducted on a subset of the available videos (the learning set) so as to define a prediction function taking features as input and outputting a fertility score. The scores computed by this function were compared to scores assigned by experts. We used the coefficient of determination (commonly denoted by R^2) as a performance measure of the learned prediction function. The best results were obtained using tracking on the ram videos showing fluorescent beads with a 4x zoom ($R^2 = 0.9$). The results on the goat videos were not as satisfying ($R^2 \simeq 0.65$), but the partner providing the videos expressed some concerns about the quality of the acquisition campaign for these data.

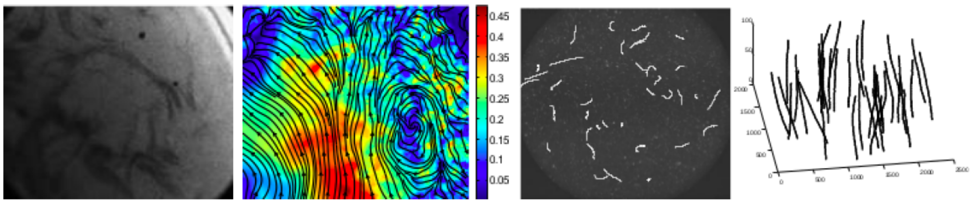


Figure 10. Predicting fertility based on massal motility measures. (Left) One frame of a video for the "drop" modality; (Middle) An example of computed optic flow; (Right: image+plot) An example of bead trajectories.

5.13. Sample selection for SVM learning on large data sets

Participants: Sonia Chaibi, Xavier Descombes, Eric Debreuve.

Support Vector Machines (SVM) represent a popular framework of supervised learning. However, it is not well adapted to large data sets since learning is performed by an optimization procedure involving the whole data set. Yet, in the end, only a small subset of the samples (the so-called support vectors) is retained for prediction. Of course, efficient algorithms exist. Still, it can be interesting to filter out as many samples as possible (the ones that will surely not be part of the support vectors) before initiating the learning procedure.

Sonia Chaibi, a PhD student from UBMA, Algeria, visited the team for a month to collaborate on this subject. The method relies on successive unsupervised sample clustering steps. After each clustering, the homogeneity of the clusters in terms of sample class assignment is used to decide which samples are unlikely to be close to the separation hyperplane (and hence unlikely to be selected as support vectors), and which samples are apparently close to this hyperplane. The former ones can be discarded, thus reducing greatly the number of samples to be processed by the SVM algorithm, while the latter ones are kept, preserving the precision of the separation hyperplane as much as possible.

5.14. Morphological Analysis and Feature Extraction of Neurons from Mouse Cortices Multiscale 3D Microscopic Images

Participants: Alexis Zubiolo, Xavier Descombes, Eric Debreuve.

This work is jointly conducted with Kawssar Harb and Michèle Studer (iBV).

We propose a framework to analyze the morphology of mouse neurons in the layer V of the cortex from 3D microscopic images. We are given 8 sets of images, each of which is composed of a 10x image showing the whole neurons, and a few (2 to 5) 40x images focusing on the somas. The framework consists in segmenting the neurons on both types of images to compute a set of specific morphological features, and in establishing the correspondence between the neurons to combine the features we obtained, in a fully automatic fashion. On the 10x images, we use a multiple birth and cut algorithm to segment the sections of the apical dendrites. Merging these intersections provides the localization of the first branching of the apical dendrite (see Fig. 11 (left)). On the 40x images, we compute an hysteresis threshold to obtain a first segmentation (somas and dendrites starts) and apply iterative morphological operators to reconstruct the full dendrites (see Fig. 11 (middle)). The correspondence map between the two types of images is done using a bipartite graph matching model that associates each neuron configuration of a 40x image – a constellation – to a subset of neurons in the 10 image – the galaxy – (see Fig. 11 (right)).

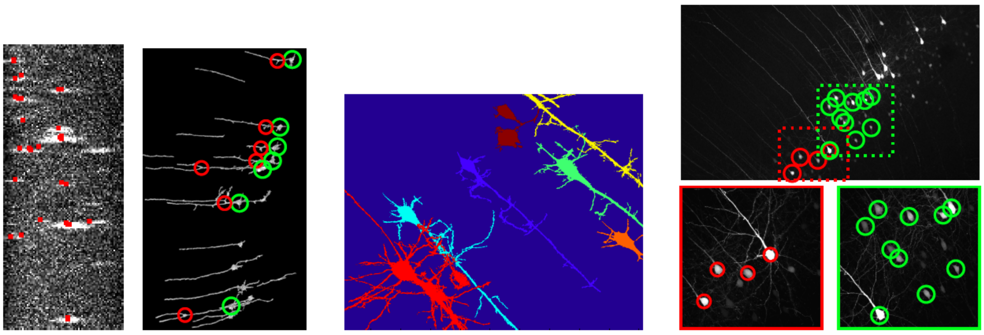


Figure 11. Left: neuron segmentation on the 10x image. Middle: full neuron reconstruction from the 40x image. Right: 10x-40x image maching.

5.15. Whole-Slide Image Analysis of Renal Cell Carcinoma

Participants: Ana Rita Lopes Simoes, Eric Debreuve, Alexis Zubiolo, Xavier Descombes.

This work is jointly conducted with Thierry Pourcher, and Philippe Pognonec (TIRO, CEA-CAL-UNSA), and Damien Ambrosetti (CHU, Nice).

We study histology images of kidney cancer that present different subpopulations of cells (tumor, healthy tissue, stroma, fat, blood, ...). The goal is to analyze the images to help determine the cancer type and stage. Given the resolution of the images ($0.25\mu m$) that leads to very large images (around 100k×100k pixels), a multiscale approach has been considered. At a larger scale, we focus on the cellular architecture and the vascular networks. Regions of interest (ROIs) have been detected with a pixelwise clustering based on neighborhood features (see Fig. 12 (left)). At a smaller scale, we extract more precise information from the cells (nucleus and cytoplasm sizes, shapes and colors, ...). The nuclei of the cells have been segmented using an Hessian determinant-based method (see Fig. 12 (middle)) which enables us to establish statistics about their size. Information on the vascular arborization has been extracted with a Frangi vesselness followed by a cleaning and gap filling post-processing (see Fig. 12 (right)).

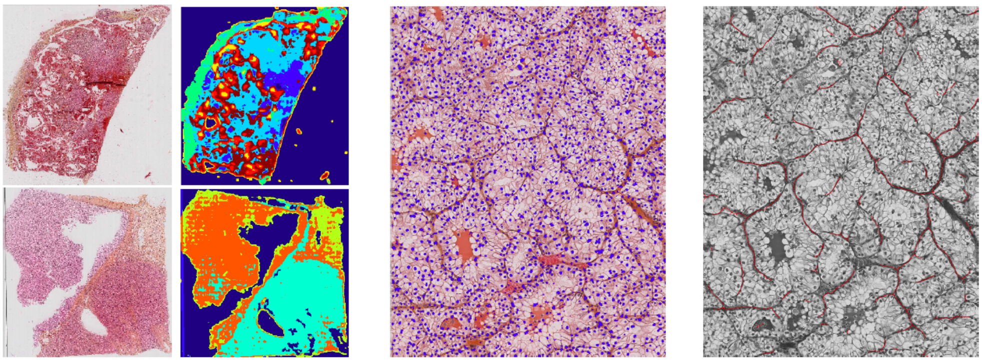


Figure 12. ROI detection (left), Nuclei segmentation in blue (middle) and Vascular arborization extraction in red (right)

SERPICO Project-Team

6. New Results

6.1. Patch-based statistical denoising methods for electron and light microscopy

Participants: Charles Kervrann, Frédéric Lavancier.

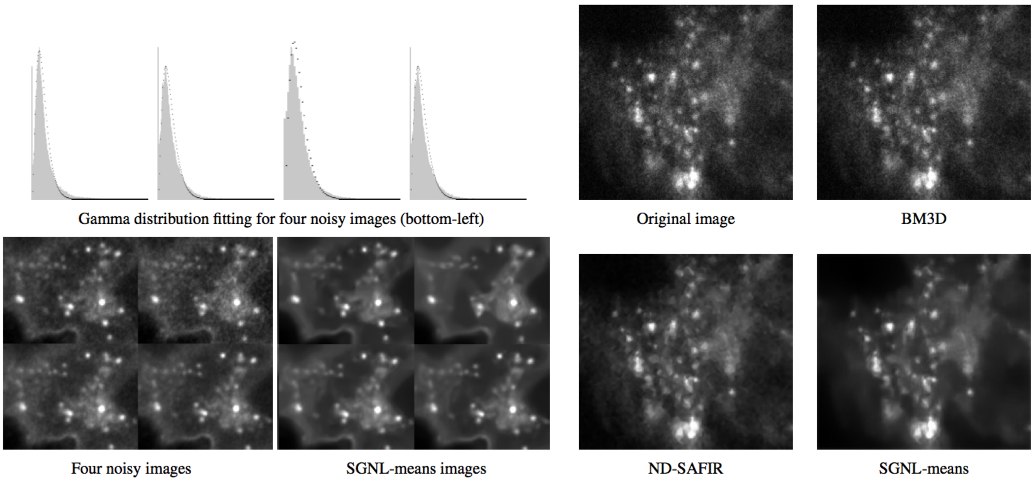
Inspired form the non-local means [33], we developed a stochastic NL-means-based denoising algorithm for generalized non-parametric noise models [21], [9]. First, we provided a statistical interpretation to current patch-based neighborhood filters and justify the Bayesian inference that needs to explicitly account for discrepancies between the model and the data. Furthermore, we investigated the Approximate Bayesian Computation (ABC) rejection method [30], [46] combined with density learning techniques for handling situations where the posterior is intractable or too prohibitive to calculate. This is particularly relevant for images contaminated by heterogeneous sources of noise. A major difference with previous methods is that we directly handle the structure of the noise, without precise parametric modeling of the noise. We demonstrated the flexibility of our stochastic Gamma non-local means (SGNL-means) by showing how it can be adapted to tackle noise in frequency domain fluorescence lifetime imaging microscopy (FD-FLIM) and cryo-electron tomography (see Fig. 3).
Moreover, we also proposed a general statistical aggregation method which combines image patches denoised with several commonly-used algorithms [20]. We showed that weakly denoised versions of the input image obtained with standard methods, can serve to compute an efficient patch-based aggregated estimator. In our approach, we evaluate the Stein's Unbiased Risk Estimator (SURE) of each denoised candidate image patch and use this information to compute the exponential weighted aggregation (EWA) estimator. The aggregation method is flexible enough to combine any standard denoising algorithm and has an interpretation with Gibbs distribution. The denoising algorithm (PEWA) is based on an MCMC sampling and is able to produce results that are comparable to the state of the art ([4], [38]). In this range of work, we have also introduced in [28] a general method to combine estimators in order to produce a better estimate. From a theoretical point of view, we proved that this method is optimal in some sense. It is illustrated on standard statistical problems in parametric and semi-parametric models where the averaging estimator outperforms the initial estimators in most cases. As part of an on-going work, we are applying this method to improve patch-based image denoising algorithms.

References: [9] [21] [20] [28]

Collaborators: Philippe Roudot (UT Southwestern Medical Center, Dallas (TX))
Francois Waharte (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Paul Rochet (Laboratoire de Mathématiques Jean Leray (LMJL), university of Nantes)

6.2. Design of deconvolution algorithms for low exposure fluorescence microscopy images

Participants: Deepak George Skariah, Charles Kervrann.



Gamma distribution fitting for four noisy images (bottom-left)

Original image

BM3D

Four noisy images

SGNL-means images

ND-SAFIR

SGNL-means

Figure 3. Experiments in FD-FLIM (confocal spinning-disk microscopy, UMR 144 CNRS-Institut Curie, PICT-IBiSA). Left: FNAR1 tagged with Green Fluorescence Protein (GFP) observed in a epithelial cell with mCHerry-tagged Tyk2 ; Gamma distribution fitting and SGNL-means denoising on four successive images with temporally varying signal-to-noise ratios. Right: comparison of denoised images with methods [4] [38].

Fluorescence imaging is popular in cell biology research due to its high contrast imaging capability. In microscopy imaging under low exposure conditions, the image quality is limited by out-of-focus blur and high noise. As a result a preprocessing stage known as deconvolution is needed to estimate a good quality version of the observed image. We proposed to design an efficient deconvolution algorithm for fluorescence microscopy under low exposure conditions by using the Poisson noise model. The result of deconvolution depends heavily on the choice of the regularization term. The regularization functional should be designed to remove noise while retaining the image structure. The choice of Poisson noise model and new regularization functional demands the design of a new and efficient optimization algorithm. We proposed to use a complex non quadratic regularization functional along with Poisson noise assumption for the first time. The use of non quadratic regularization makes the resulting optimization problem a complex one. This demanded the development of a problem-specific optimization algorithm which is fast as well as robust enough to minimize a non quadratic cost function. The use of non quadratic regularization together with Poisson noise model ensures that finer details of underlying structures are well restored in the presence of large amount of noise.

Collaborator: Muthuvel Arigovindan (Imaging Systems Lab, Department of Electrical Engineering, Indian Institute of Science, Bangalore, India).

6.3. Background estimation and vesicle segmentation in live cell imaging

Participants: Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

In live cell fluorescence microscopy images, the moving tagged structures of interest, such as vesicles, often appear as bright spots with intensity that varies along time over a time-varying and cluttered background. Localization and morphology assessment of these small objects over time is then crucial to provide valuable information for quantitative traffic analysis. In this study, we have focused on the Rab6 protein as a typical intracellular membrane-associated protein. Rab6 is known to promote vesicle trafficking from Golgi to Endoplasmic Reticulum or to plasma membrane. In our study, micro-fabricated patterns have been used to enforce cells to have circular or crossbow normalized shape. Micro-patterns impose constraints on the cytoskeleton and the location of organelles (e.g. Golgi apparatus) is thus better controlled. These micro-patterns also influence the spatial distribution of Rab6 transport carriers. However, the direct influence of the micro-patterns on the spatial dissemination of these trafficking vesicles has so far not been completely characterized. In this work, we have considered a statistical Bayesian approach in the framework of conditional random fields (CRF) for background estimation and vesicle segmentation [13]. Within this approach, we have designed a robust detection measure for fluorescence microscopy based on the distribution of neighbor patch similarity. We formulate the vesicle segmentation and background estimation as a global energy minimization problem. An iterative scheme to jointly segment vesicles and background is proposed for 2D-3D fluorescence image sequences. We have conducted a quantitative comparison with state-of-the-art methods on a large set of synthetic image sequences with a cluttered time-varying background and achieved a quantitative validation of the vesicle segmentation method on 2D and 3D micro-patterned cells expressing GFP-Rab6.

Reference: [13]

Collaborators: Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

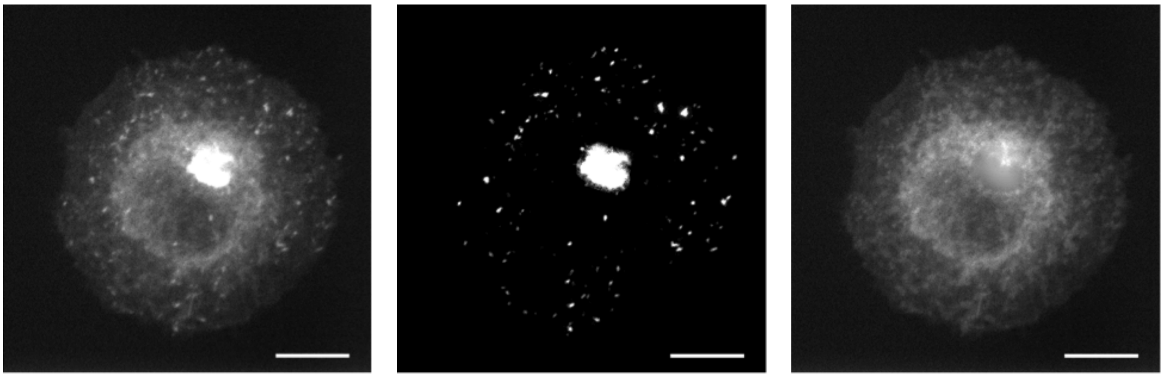


Figure 4. Left: Fuorescence confocal spinning-disk microscopy image depicting GFP-Rab6 proteins (UMR 144 CNRS-Institut Curie, PICT-IBiSA). Middle: estimated vesicular component. Right: estimated background.

6.4. A quantitative approach for space-time membrane trafficking orientation

Participants: Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

Rab6 proteins are trafficking from the Golgi apparatus at the cell center to Endoplasmic Reticulum or to plasma membrane located at the periphery of the cell. The cell shape influences Rab6 trafficking but no study has ever quantified the effect of the cell shape on the trafficking orientation. In this study [25], we compare Rab6 trafficking orientation constrained by two different micropatterns [56] (circular and crossbow-shaped cells) from fluorescence video-microscopy. Object/background separation [13] is first applied to 3D+T image sequences to extract Rab6 spatio-temporal coordinates. The bandwidth of the von Mises kernel is automatically estimated using the rule of thumb and leads to two different densities for the two different micropatterns. We propose to quantitatively compare these densities by computing the Wilcoxon rank sum paired test between inter- and intra-micropattern distances. We considered the circular earth mover's distance (also known as the Wasserstein metric) to compare traffic densities. Our quantitative study on micro-patterned cells concludes that the Rab6 transport carriers destinations concentrate at the three corner points of the crossbow-shaped cells corresponding to the main adhesion sites, while the vesicle destination distribution is somewhat uniform for circular-shaped cells.

Reference: [25]

Collaborators: Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

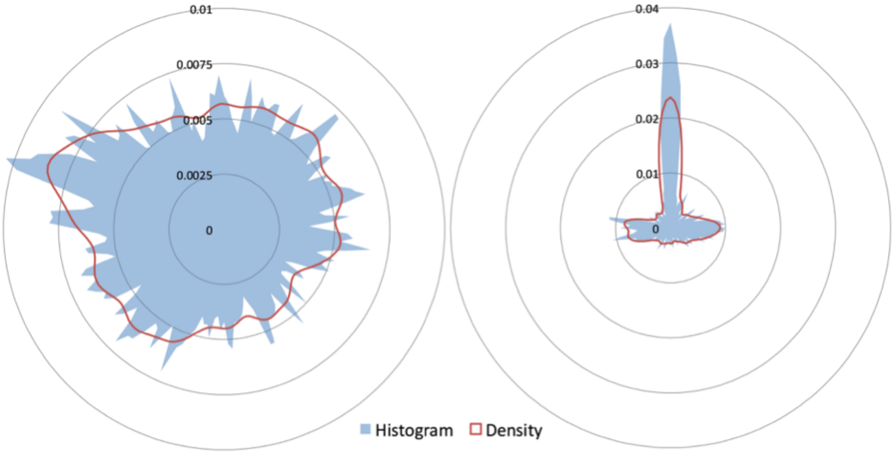


Figure 5. Distribution of traffic orientation for circle-shaped cells (left) and crossbow-shaped cells (right).

6.5. Vesicle segmentation method with automatic scale selection in TIRF microscopy

Participants: Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Accurately detecting subcellular particles in fluorescence microscopy is of primary interest for further quantitative analyses such as counting, tracking or classification. Our primary goal was to segment vesicles in fluorescence microscopy images. In [15] we proposed a first spot detection method with automatic scale selection. We have now dramatically improved the precision of the scale selection step, yielding to a more reliable detection of the spots [23]. The method relies on a Laplacian of Gaussian (LoG) filter to first enhance the spots while reducing noise. To obtain good detection results, the scale of the Gaussian filter must be precisely set, according to the spots size [23]. In order to cope with very small spots, we rely on the discrete analog of the Gaussian filter [45], instead of the previously used sampled Gaussian filter. With this filter, we can find the optimal Gaussian scale with an arbitrary precision by minimizing a statistical criterion. We have introduced two criteria for this purpose and compared them. Once the optimal scale is selected, we threshold the lowest values of the LoG-filtered image, which correspond to spots. To cope with inhomogeneous background, thresholding must be adapted to local statistics so that a single probability of false alarm (PFA) setting can be defined for the whole image or even the collection of images to be processed. In short, we automatically infer from image data the optimal parameters usually left to the user guidance in other methods, that is, spot scale and detection threshold. We have carried out an extensive comparative evaluation, which demonstrates that our new scale selection approach improves detection performances, and that our spot detection method outperforms state-of-the-art detectors [23].

References: [15] [23]

Collaborators: Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

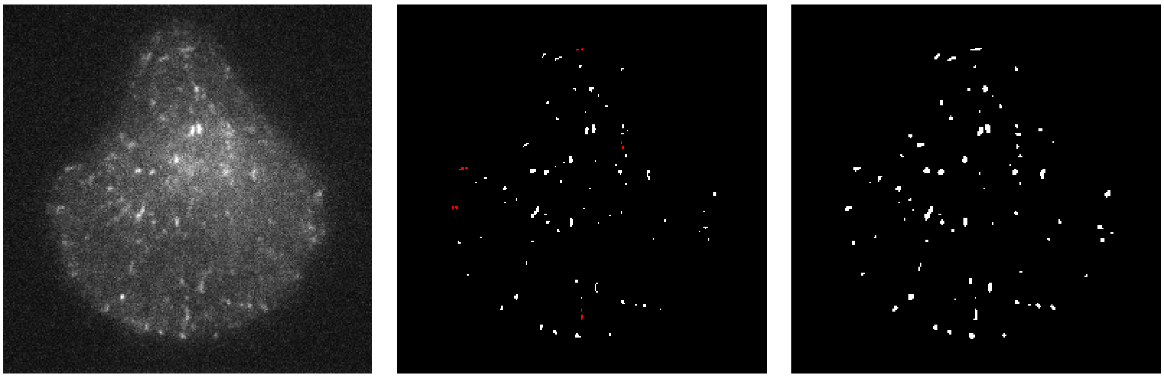


Figure 6. Comparison of segmentation results on a real image presenting elongated spots. Left: Input TIRFM images (Rab11-mCherry) (UMR 144 CNRS-Institut Curie, PICT-IBiSA). Middle: Segmentation results with state-of-the-art detector MS-VST [60]. Some elongated spots of (left) are split (red) by MS-VST due to a too small filter scale. Right: Segmentation results with our new detection method. Elongated objects are well recovered thanks to the precise scale selection.

6.6. Analysis of the repartition of moving vesicles by spatio-temporal point process models

Participants: Frédéric Lavancier, Thierry Pécot, Charles Kervrann.

Characterizing the spatial repartition of interacting moving proteins is a fundamental step for co-localization and co-expression. Based on the segmentation algorithm [15], [23], this challenge amounts to characterizing the repartition or spatial distribution of spots (see Fig. 6). This is part of the more general statistical analysis of random geometrical objects, and in particular of random points. Gibbs models form a large class of point process models, that can be used to characterize either complete randomness or attraction or repulsion between points depending on the Gibbs potential at hand.

First in [27], we focused on infinite range potentials that include the most famous interaction potential arising from statistical physics, namely the Lennard Jones potential. To fit this kind of models to a dataset, the standard inference methods are not applicable. We introduced in [27] a modification of the pseudolikelihood method, with a specific border correction, and we prove that this provides consistent and asymptotically normal estimators. Second, in [26], we studied an alternative class of models, the determinantal point processes (DPP). They are designed to model repulsion between points and are thus adapted to regular point patterns. These models are becoming very popular in the spatial statistics community due to many appealing properties. We quantified the possible repulsiveness that a DPP can model [26]. In particular, we determined the most repulsive stationary DPP. We finally introduced new parametric families of DPPs that cover a large range of DPPs, from the homogeneous Poisson process (for no interaction) to the most repulsive DPP.

An application of these models to the problem of co-localization between proteins is part of an on-going project. In each protein, the set of vesicles is modeled by a union of random balls, possibly overlapping, and a Gibbs interaction is introduced to take into account the possible interaction in the location of vesicles between two proteins. Our first concern is to test whether the two proteins actually interact, i.e. co-localization occurs, or in other words whether the Gibbs interaction is empty or not. If there is co-localization, the further step is to characterize it through the estimation of the strength of the Gibbs interaction.

References: [26] [27]

Collaborators: Christophe Ange Napoléon Biscio (LMJL, University of Nantes)

Jean-François Coeurjolly (Laboratoire Jean Kuntzmann, Grenoble Alpes University)

6.7. Detection and estimation of membrane diffusion during exocytosis in TIRF microscopy

Participants: Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Assessing the dynamics of plasma membrane diffusion processes in live cell fluorescence microscopy is of paramount interest to understand cell mechanisms. We investigated methods to detect vesicle fusion events, and estimate the associated diffusion coefficients in TIRFM image sequences [16]. In contrast to classical approaches, a diffusion coefficient is locally estimated for each detected fusing vesicle. We first detect the membrane fusion events and then select the diffusion configurations among them with a correlation test. To estimate the diffusion coefficient, a geometric model is fitted to the detected spot directly in the 2D+T subvolume. This recent estimation approach produced more satisfying results when compared to [16]. Diffusion events are reliably recognized, and the diffusion coefficient is accurately estimated for each diffusion event. This work will be integrated in a broader study, spanning from transport phase to membrane fusion, and non-diffusion events will be analyzed.

Reference: [16]

Collaborators: Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

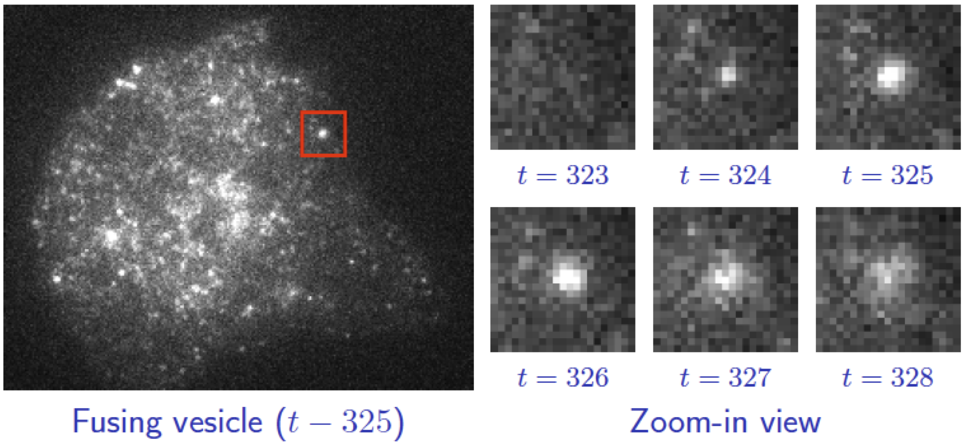


Figure 7. Left: Fusing vesicle (frame in red) in a TIRFM (UMR 144 CNRS-Institut Curie, PICT-IBiSA) sequence (frame 325, 50ms/frame). Right: Zoom-in view of the temporal evolution of the fusing vesicle.

6.8. Estimation of the flow of particles without tracking in fluorescence video-microscopy

Participants: Thierry Pécot, Patrick Bouthemy, Charles Kervrann.

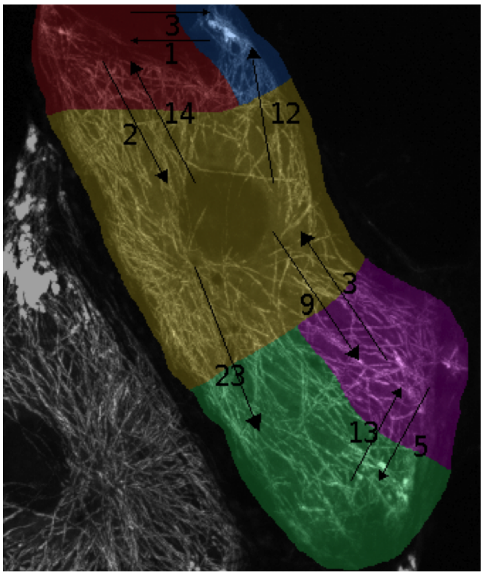


Figure 8. Vesicle flows estimated when considering a simple partition of 5 regions for an image sequence acquired in TIRF microscopy and showing the protein Clip170 (UMR 144 CNRS-Institut Curie, PICT-IBiSA).

Automatic analysis of the dynamic content in fluorescence video-microscopy is crucial for understanding molecular mechanisms involved in cell functions. We have proposed an original approach for analyzing particle trafficking in these sequences. Instead of individually tracking every particle, we only locally count particles crossing boarders between regions over time and minimize a global energy function. Three methods to determine the particle flow have been considered. We have conducted comparative experiments on synthetic and real fluorescence image sequences. We have shown that adding a sparsity constraint on the number of detected events allows us to reduce the number of false alarms. Compared to usual tracking methods, our approach is simpler and the results are very stable. This estimation method needs the adjustment of only two parameters. (see Fig. 8).

Reference: [22]

Collaborators: Jean Salamero (UMR 144 CNRS-Institut Curie, STED team and PICT-IBiSA)
Jérôme Boulanger (UMR 144 CNRS-Institut Curie, STED team)

6.9. Detection and tracking of astral microtubules at the cell cortex

Participants: Thierry Pécot, Charles Kervrann, Geoffrey Dieffenbach.

In this study, we are interested in the influence of the mechanical properties of astral microtubules in the centering mechanisms of the mitotic spindle, giving it a robust positioning. In their previous studies, the CeDRE group (IGDR Rennes) identified two subpopulations of astral microtubules that either push or pull the cell cortex. To better understand these mechanisms, image sequences are acquired at the cortex level where extremities of astral microtubules come to exert forces. In order to characterize the two subpopulations of astral microtubules during the mitosis in the unicellular embryos of C. Elegans, life span, that is the period during which the microtubule is touching the cell cortex, for every single microtubule has to be measured. A short life span corresponds to a pulling force while a longer life span corresponds to a pushing force. Detecting and tracking microtubules at the cell cortex has to be done to collect these measures. As the signal-to-noise ratio is low, a denoising step is needed to detect the microtubule extremities. Several detection methods were tested but we need to further investigate this step to find the most suited methods for this particular application. Finally, the U-track algorithm [42] is applied to track the microtubules extremities to measure their life span.

Collaborators: Jacques Pécréaux (CeDRE group, IGDR Rennes, CNRS UMR 6290)
Hélène Bouvrais (CeDRE group, IGDR Rennes, CNRS UMR 6290)

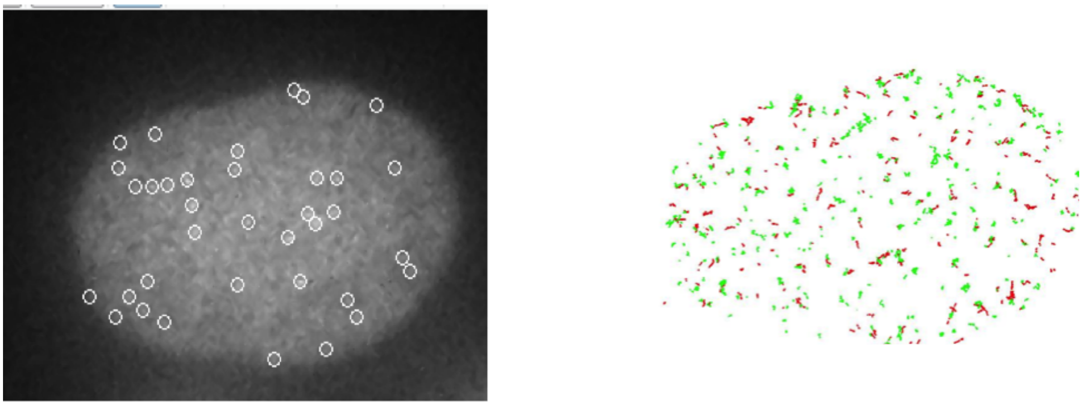


Figure 9.

Microtubule extremities detection and tracking in fluorescence microscopy
(embryo of C. Elegans, IGDR - Institute of Genetics and Developmental biology of Rennes, CNRS UMR 6290).

6.10. Spot localization and segmentation for Tissue MicroArray (TMA) de-arrying

Participants: Hoai Nam Nguyen, Charles Kervrann.

Tissue core de-arraying is one of the most important steps in tissue microarray (TMA) image analysis. A very first task of TMA (Tissue MicroArray) image analysis is to accurately localize spots (separate tissue core) representing arrays of 512×512 pixels each, in very large images of several thousands of pixels. However, few solutions and frameworks are available and none of them covers images provided by fluorescent scanners. We developed a robust TMA de-arraying method adapted for digital images from classical optical and new fluorescent devices. The proposed algorithm is composed of three modules: i) detection, ii) segmentation, and iii) array indexing. The detection of TMA cores is performed by local adaptive thresholding of isotropic wavelet transform coefficients. We demonstrated how a wavelet decomposition at any desired scale can be performed faster than usual techniques by exploiting explicit formula of the analysis wavelet. Our core detection strategy enables to deal with images having significant noise level, inhomogeneous background, and high dynamic range such as fluorescence images, without any assumption on image noise and intensity value range. The detected cores are furhter segmented by using parametric ellipse model to improve detection accuracy. Combining these two modules, we can handle complex background and artifacts, particularly in fluorescence imaging, and thus reduce false detections. After the segmentation step, the position of detected cores is determined by the centroid of relevant segments. Finally, to compute array indices of cores, we estimate the deformation of a theoretical grid under a thin-plate model by using an iterative scheme. After each iteration, the initial regular grid is progressively transformed for fitting computed core positions. Our main contribution is the reformulation of the array indexing problem as an estimation of the deformation function, which is solved with a iterative algorithm. Moreover, when design layout of TMA slide is known, our estimator of deformation yields quantitative information about grid deformation such as average translation, rotation angle, shearing coefficients, bending energies along axis, etc. They can be used as quality indicators of the manufactured TMA slide.

Collaborator: Vincent Paveau (Innopys company)

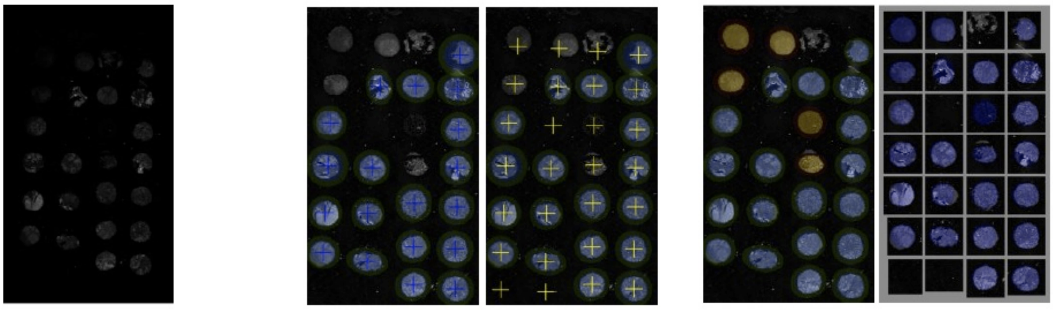


Figure 10. Array indexing TMA (Innopsys company). From left to right : input TMA image, segmented core positions marked by blue crosses, estimated positions of deformed grid marked by yellow crosses, retrieved missed cores after detection/segmentation steps (orange areas), and array representation of TMA (retrieved cores are colored).

6.11. Adaptive global and local motion estimation

Participants: Noémie Debroux, Charles Kervrann.

The design of data costs is one of the main research issue for variational optical flow estimation. The aim is to improve discriminative power by integrating appropriate neighborhood information, while preserving computational efficiency. Most previous works define features on patches with predefined sizes and shapes, or filter pixelwise costs with fixed filtering parameters. We proposed a novel approach estimating spatially varying parameters of filters used to define the data term [8]. More specifically, our model considers Gaussian filtering of the pixelwise brightness constancy equation and imposes smoothness constraints on motion and convolution filter size (bandwith). The energy encoding these assumptions is alternatively minimized over flow field and the spatially varying bandwidth in a variational framework. Experimental results on the Middlebury database demonstrated clear improvements yielded by our method over the spatially constant case of [32] (see Fig. 11).

Collaborator: Denis Fortun (UMR 144 CNRS-Institut Curie, STED team, Paris)
(EPFL, Lausanne, Switzerland)

6.12. Crowd motion classification

Participants: Antoine Basset, Charles Kervrann, Patrick Bouthemy.

Assessing crowd behaviors from videos is a difficult task while of interest in many applications. We have defined a novel approach which identifies from two successive frames only, crowd behaviors expressed by simple image motion patterns. It relies on the estimation of a collection of sub-affine motion models in the image, a local motion classification based on a penalized likelihood criterion, and a regularization stage involving inhibition and reinforcement factors [17]. The apparent motion in the image of a group of people is assumed to be locally represented by one of the three following motion types: translation, scaling or rotation. The three motion models are computed in a collection of predefined windows with the robust estimation method [48]. At every point, the right motion model is selected owing to the corrected (for small sample size) Akaike information criterion (AICc). To classify the local motion type, the three motion models are further subdivided into a total of eight crowd motion classes. Indeed, scaling refers either to gathering (Convergence) or dispersing people (Divergence). Rotation can be either Clockwise or Counterclockwise.

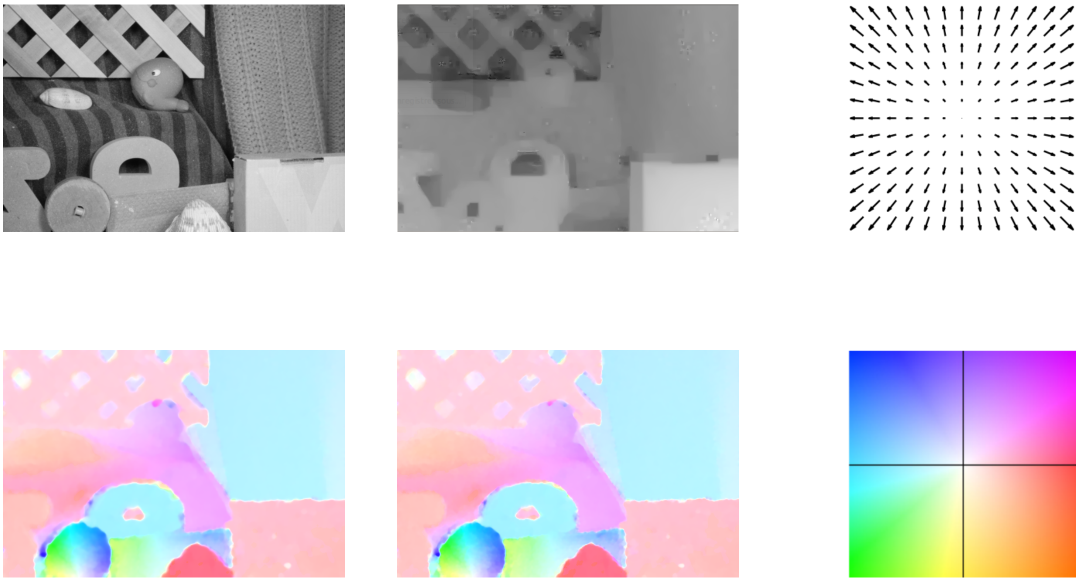


Figure 11. Comparison on a sequence of the Middlebury benchmark. Top from left to righ: input image and spatially filter bandwidth estimation. Bottom from left to right: velocity field computed by [32] (endpoint error = 0.143) and by our method (endpoint error = 0.126).

Since our classification scheme is view-based, four image-related translation directions are distinguished: North, West, South, East. Then, to get the final crowd classification, a regularization step is performed, based on a decision tree and involving inhibition for opposed classes such as convergence and divergence. We have also developed an original and simple method for recovering the dominant paths followed by people in the observed scene. It involves the introduction of local paths determined from the space-time average of the parametric motion subfields selected in image blocks. Starting from one given block in the image, we straightforwardly reconstruct a global path by concatenating the local paths from block to block. Experiments on synthetic and real scenes have demonstrated the performance of our method, both for motion classification and principal paths recovery.

Reference: [17]

6.13. Anomaly detection using block-based histograms of crowd motion patterns

Participants: Juan Perez Rua, Antoine Basset, Patrick Bouthemy.

We have developed a new and generic method to detect and localize abnormal events in videos of crowd scenes. The algorithm consists first in determining the flow vector and crowd motion class for every moving pixel from a set of affine motion models estimated on a collection of windows. Then, the observed scene is subdivided in blocks to compute crowd motion class histograms weighted by the motion vector magnitudes. A very simple training step enables to get the reference histograms per block accounting for the normal behaviours. For each block, we can automatically set by means of statistical arguments the threshold on the distance between the histogram in the current image and the reference histogram that decides the presence of an abnormal event in that block. Results of extensive experimentation on different types of anomaly datasets show that our method is competitive with respect to methods relying on far more elaborated models on both appearance and motion and thus involving a significant learning stage. It outperforms any other existing purely motion-based anomaly localization method.

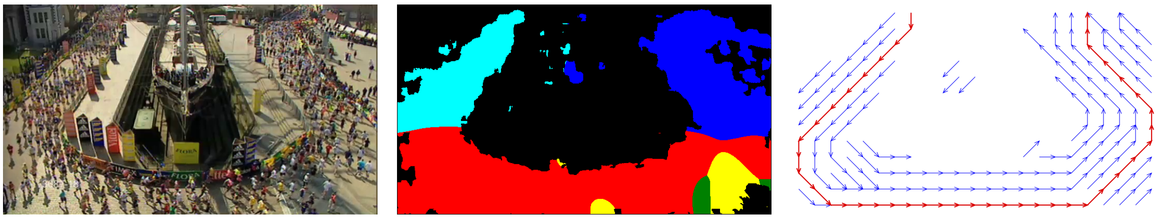


Figure 12. Overview of the method applied to a sequence where runners follow a ‘U’ from the upper left corner to the upper right corner. Left: First frame of the sequence. Middle: Classification results (cyan=translation toward South, red=counterclockwise rotation, yellow=translation toward East, green=convergence, blue=translation toward North). Right: Recovery of the longest path in the scene (red).

VIRTUAL PLANTS Project-Team

5. New Results

5.1. Highlights of the Year

- Publication of a joint work with RDP at ENS-Lyon in the journal 'Nature'. In December 2013, a joint work on phyllotaxy with the RDP lab from ENS-Lyon was published online in the journal Nature [2]. This paper obtained the 2014 prize "la Recherche" in the biology category http://www.leprixlarecherche.com. Based on the analysis of phyllotaxis perturbations in mutants, this study sheds a new light on our interpretation of phyllotaxis, revisiting the standard model and suggesting that several fields based on auxin and cytokinin with different properties are required to provide robustness to phyllotaxis.
- To study and model morphogenesis, the team has been working in the last 8 years on modeling mechanical forces and deformations in tissues in collaboration with the UMR RDP at ENS-Lyon. This work has given rise to the development of a 3D computational framework to model the mechanics of 3D plant tissues during growth at cellular resolution and has been finalized this year with a publication in PLoS Comp. Biology (to appear in 2015). This framework makes it possible to construct models of meristem development, showing how the regulation of regional identities can lead to realistic shape development by dynamically modulating the mechanical properties of cells. It has been used also to study the influence of a specific signalling cascade (the ABP1-Kat1 signalling pathway) and its putative mechanical consequences on primordium initiation [25]. The expertize gained by our groups on physical models of plant tissue development has been wrapped up in a review paper [12].

5.2. Analysis of structures resulting from meristem activity

5.2.1. Acquisition and design of plant geometry

Participants: Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier, Ibrahim Chedaddi, Mathilde Balduzzi, Julien Diener.

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (ligth, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [61], [52]. These methods are based on direct measurements of position and shape of every plant organ in space. Although they provide accurate results, they are particularly time consuming. More rapid and automated methods are now required in order to collect plant architecture data of various types and sizes in a systematic way. In this aim, we explore the use of pictures, laser scanner and video.

- *Reconstruction of plant architecture from 3D laser scanner data.* (Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Godin, Pascal Ferraro [Labri, Bordeaux], Yassin Refahi)

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developing algorithms to reconstruct branching systems without leaves or foliage from scanner data or from scan simulated on plant mock-up obtained using different digitizing method.

For branching systems, a number of automatic methods were proposed in the literature to reconstruct plausible branching structures from laser scanner data. The question of their assessment and accuracy is however critical for further exploitation in biological applications. For this, we developed an evaluation pipeline to assess reconstruction accuracy. A laser scan database on which experts built reference reconstructions is used as a basis of the evaluation. A graphical editor has been developed to help experts to reconstruct semi automatically reference structures. The evaluation pipeline is given two plant structures and compares their organization. Similar elements are identified based on geometric criteria using an optimization algorithm. The organization of these elements is then compared and their similarity is quantified. Two indices of geometrical and structural similarities are defined, and automatic reconstructions can thus be compared to reference structures to assess their accuracy. The method is successful at capturing the variation of similarities between two structures as different levels of noise are introduced. A first comparative evaluation of the different methods of the literature has been designed and conducted. This work has been published in the special issue of Annals of Botany on FSPMs [16]. A procedure to automatically determine phyllotactic angles from scans of small plants has been added recently to the reconstruction pipeline.

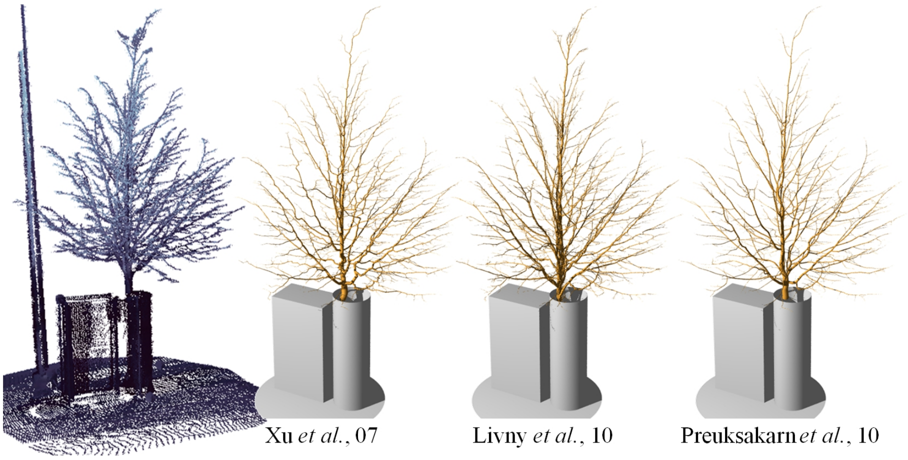


Figure 2. Comparison of different methods of the litterature to reconstruct plant architecture from laser scanner data. On the left, the original scan and on the rigth, the resulting reconstructions using three different methods. These reconstructions are quantitativelly assessed using our evaluation pipeline presented in [16]

In the context of the PhD of M. Balduzzi, we also investigated the reconstruction of tree foliage from 3D scans. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers in the silhouette of the scans that make the meshing of the pointset extremely difficult. New generation of laser scanners provide intensity of the laser reflected on the surface of scanned objects. This intensity depends on the distance to the object, its optical property and the incidence angle. A first work on this topic showed that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we developed a reconstruction pipeline using the scan intensities and based on Shape-From-Shading. Outliers being along the edge of the surface point cloud, we chose to develop a propagation SFS method initialized with points of the scans with high quality. We proved that surface with constant intensity are necessarily surfaces of constant slope or sand-pile surfaces. Using this result, a propagation method along iso-intensity regions was developed. These surfaces can then be sampled to provide a smooth point set without outliers. This work has been presented by M. Balduzzi for her thesis defense in november.

- *Reconstruction from video.* (Frédéric Boudon, Jerome Guenard [IRIT, Toulouse], Geraldine Morin [IRIT, Toulouse], Pierre Gurdjos [IRIT, Toulouse], Vincent Charvillat [IRIT, Toulouse])

  Even if mature computer vision techniques allow the reconstruction of challenging 3D objects from images, dedicated methods for generating 3D plant models must be devised due to the high complexity of plant topology. In collaboration with our colleagues from IRIT, Toulouse, we developed an analysis-by-synthesis method which generates 3D models of plants from both images and a priori knowledge of the plant species.

  Our method is based on a skeletonisation algorithm which allows to generate a possible skeleton from a foliage segmentation. Then, a 3D generative model, based on a parametric model of branching systems that takes into account botanical knowledge is built. This method extends previous works by constraining the resulting skeleton to follow hierarchical organization of natural branching structure. 3D models are then generated. 2D projections of the models can be compared with the original image to assess the visual accuracy of the reconstruction. We optimise the parameter values of the generative model based on the 2D projection criterion. Realistic results are obtained on different species of plants, in particular vineyards. This work has been presented at the Mathematical Methods for Curves and Surfaces conference and published in LNCS [42].

- *Reconstruction of virtual fruits from pictures.* (Ibrahim Chedaddi, Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2])

  This research theme is supported by the Agropolis project MecaFruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function. To this end, we used a modeling pipeline previously developped by Mik Cieslak in the OpenAlea platform. This methods involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex network of vasculature that is embedded within this mesh. Previous studies demonstrated the possibility to create species-specific models of fruit structure with relatively low effort [49]. We focus now on validating the vasculature networks by quantitatively comparing them to experimental data from the litterature.

  These physiological data will be combined with a mechanical model of fruit growth, to investigate the effects of fruit structure on quality (see section 5.4.2).

- *Reconstruction of root structures.* (Julien Diener, Fredéric Boudon, Christophe Pradal, Christophe Godin, Philippe Nacry [BPMP, INRA], Christophe Périn [AGAP, CIRAD], Anne Dievart [AGAP, CIRAD], Xavier Draye [UCL, Belgium])

  This research theme is supported by the Agropolis through the Rhizopolis project and by NUMEV.

  Similarly to aerial part of plants, new needs for automatic digitizing of root systems emerge. Most existing methods focus only on semi-automatic approaches. This does not support the high-throughput capabilities of acquisition systems. In the context of the RhizoScan project, we previously designed a prototype of an automatic image analysis pipeline to extract root system architecture of branching systems grown in Petri boxes. This pipeline provides i) a set of model based image segmentation method, ii) the extraction of a graph representation of the root system, and iii) a method to identify the root axes organization. This year, we improved and extended the pipeline in the following way:

  1. We integrated a validation step in the workflow based on the comparison method presented in [16].
  2. We participated in an international collaboration with the Université Catholique de Louvain (Belgium), the CPIB of the University of Notthingham (UK), the University of Vienna (Austria), the Jülich research center (Germany) and INRA, to develop a standard file format for root architecture. The resulting format (RSML) is described in a publication to appear in *Plant Physiology* in 2015 (details can be found at rootsystemml.github.io).

3. We have initiated a collaboration to integrate visual data mining methods developed by the Zenith team in order to improve the automation of the image analysis pipeline.
4. In general, the robustness of the pipeline has been improved. In particular, an optimization method has been designed to select the root axes hierarchy that respect specific botanical constraints.

5.2.2. Modeling the plant ontogenic programme

Participants: Christophe Godin, Yann Guédon, Jean-Baptiste Durand, Pierre Fernique, Christophe Pradal, Jean Peyhardi.

This research theme is supported by two PhD programmes.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g. [46], [54], [55], [51], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic programme" [59], "age state" [50] or "physiological age" [48]. All these concepts state that meristem fate changes according to position within the plant structure and during its development. Even though these changes in meristem fate are specific to each species and lead to the differentiation of axes, general rules can be highlighted [50], [48]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste [AGAP, AFEF team], Catherine Trottier [I3M], Yves Caraglio [AMAP], Pierre-Eric Lauri [AGAP, AFEF team])

  Shoot branching structures often take the form of a succession of homogeneous branching zones and have been analyzed using segmentation models such as hidden semi-Markov chains. Axillary meristem fates are influenced by local properties of the parent shoot such as for instance its growth rate or local curvature. The objective of this work, which is part of the PhD subject of Jean Peyhardi, is to develop statistical models that generalize hidden semi-Markov chains with the capability to incorporate explanatory variables that vary along the parent shoot (e.g. maximum growth rate of the leaf, surface of the leaf, length of the internode, local curvature of the parent shoot). More precisely, the simple multinomial distributions that represent the axillary productions observed in the different branching zones are replaced by multinomial generalized linear models (GLMs). Since the two classical categories of multinomial GLMs that correspond either to nominal or ordinal categorical response variables were not appropriate, we chose to develop a new family of multinomial GLMs called partitioned conditional GLMs [38] that enable to tackle hierarchically-structured categorical response variables. Typically, we need to distinguish different timing of branching (e.g. immediate shoot, one-year-delayed shoot and latent bud), different categories of offspring shoots (e.g. among one-year-delayed shoots, vegetative short shoot, vegetative long shoot and flowering shoot) and to specialize the explanatory variables for certain categories of offspring shoots (e.g. the growth of the parent shoot influence the immediate offspring shoots but not the one-year-delayed offspring shoots). The resulting integrative models are called semi-Markov switching partitioned conditional GLMs and have been applied to apple and pear tree branching structures.

- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes, Yann Guédon)

  A first study was published to characterize genetic determinisms of the alternation of flowering in apple tree progenies. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot, which corresponds to portions of stem that were grown during the same year). Two replications of each genotype were available.

Indices were proposed for early detection of alternation during the juvenile phase. They were based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. This allows to quantify alternation from the yearly numbers of inflorescences at tree scale.

- *Identifying and characterizing patterns in tree-structured data* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

  In the context of Pierre Fernique's PhD (Montpellier 2 University and CIRAD), two complementary approaches were developed for analyzing patterns in tree-structured data:

  - multitype branching processes relying on local dependency properties for analyzing motifs.
  - multiple change-point models relying on long-term dependencies for segmenting trees in homogeneous zones.

  In multitype branching processes, the plant development is viewed as a demographic process, a parent entity of a given type generating child entities of different types (e.g. vegetative and flowering entities). Formally, the botanical entity properties are summarized as a categorical state variable. The number of child entities in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious generation distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs [33]. To relax the strong constraints regarding dependencies induced by parametric distributions, mixture of graphical models were also considered [41]. Multitype branching processes were applied to the analysis of the patchiness pattern (consisting of canopies made of clumps of either vegetative or flowering growth units) in mango trees.

- *Simulating fruit tree phenology* (A.S. Briand, Frédéric Boudon, Frédéric Normand [CIRAD, HortSys, Réunion Island], Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP])

  Mango is a tropical tree characterized by strong asynchronisms within and between trees. To study more precisely the interplay between the plant structural components, we built an integrative model to simulate the plant development based on the L-system formalism and GLM to model the dependencies between events. This model has been started during the internship of A. Jestin last year and has been continued during the visit of F. Boudon at the Hortsys lab in the Réunion island. The model has been extended this year to model growth and phenology of shoots and inflorescences (internship of A.S. Briand). For this, the sizes of the different organs is modelled by statistical laws estimated from measurements that depends on their positions in the architecture. The growth speed of organs is modulated by the temperature. This model will serve as a basis for further ecophysiological study in silico.

- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand [CIRAD, HortSys, Réunion Island])

  Plant growth, i.e. the increase of organ dimensions over time, and development, i.e. the change in plant structure, are often studied as two separate processes. However, there is structural and functional evidence that these two processes are strongly related. Our aim was to investigate the co-ordination between growth and development using mango trees, which have well-defined developmental stages. Developmental stages, determined in an expert way, and organ sizes, determined from objective measurements, were collected during the vegetative growth and flowering phases of two cultivars of mango. For a given cultivar and growth unit type (either vegetative or flowering), a multi-stage model based on absolute growth rate sequences deduced from the measurements was first built, and then growth stages deduced from the model were compared with hand-annotated developmental stages. Strong matches were obtained between both stages, leading to a consistent definition of integrative developmental growth stages. The growth stages highlighted growth asynchronisms between two topologically connected organs, namely the vegetative axis and its leaves. Integrative developmental growth stages emphasize that developmental stages are closely related to organ growth rates and can be interpreted in terms of the possible physiological processes (hydraulics, biomechanics and carbohydrate partitioning) underlying these stages.

- *Characterizing the successive flowering phases of strawberry in relation to genetic determinants* (Yann Guédon, Béatrice Denoyes [INRA, UMR BFP, Villenave d'Ornon], Justine Perrotte)

  Our aim was to characterize the successive flowering phases of perpetual flowering strawberry genotypes, which is of particular importance for better predicting fruit production. We applied multiple change-point models for the synchronous segmentation of the individuals of a given genotype in successive flowering phases. We identified two groups of genotypes that differ by the intensity of the flowering at the end of the flowering period. Using a genetic approach, we identified a locus controlling the flowering intensity at the end of the flowering period that likely explain these two groups of genotypes.

- *Self-nested structure of plants.* (Christophe Godin, Romain Azais, Farah Ben Naoum, Jean-Baptiste Durand, Alain Jean-Marie)

  In a previous work [6], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

  The method proposed in [6] thus compresses a tree in width, but not in height. In a new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms has been submitted recently for publication.

  The degree of self-nestedness is defined in [6] as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus designed a heuristic method based on interacting simulated annealing algorithms to tackle this difficult question. This procedure is also a keystone in a new topological clustering algorithm for trees that we propose. In addition, we obtain new theoretical results on the combinatorics of self-nested structures. The redaction of an article is currently in progress.

5.2.3. Analyzing the influence of the environment on the plant ontogenic programme

Participants: Jean-Baptiste Durand, Christian Fournier, Christophe Godin, Yann Guédon, Christophe Pradal, Jean Peyhardi, Pierre Fernique, Guillaume Garin.

This research theme is supported by three PhD programs.

The ontogenetic programme of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 5.2.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic programme from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

- *Influence of environment conditions and horticultural practices on the branching and axillary flowering structures of fruit tree shoots.* (Yann Guédon, Evelyne Costes [AFEF Team, AGAP], Ted DeJong [UC Davis], Claudia Negron [UC Davis]).

  In the context of a collaboration with Claudia Negron and Ted DeJong, we studied the influence of water availability [24] and pruning practices on the branching and axillary flowering structures of different categories of almond shoots. Stochastic models (hidden semi-Markov chains) were built for the branching and axillary flowering structures of different categories of almond shoots corresponding to different genetic backgrounds, levels of irrigation and pruning practices.

- *Analyzing growth components in trees.* (Yann Guédon, Yves Caraglio [AMAP], Olivier Taugourdeau [AMAP])

  We identified robust indicators that summarize the respective importance of ontogeny and environmental constraints (mainly related to light environment) in forest tree development. In this context, tree growth data correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models to identify tree growth phases. These segmentation models, which are hidden semi-Markov chains, were compared with simple hidden Markov chains that correspond to the environment-driven development assumption. This statistical modelling approach was applied to both evergreen (Corsican pine and silver fir) and deciduous (sessile oak and Persian walnut) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. Growth phase duration distributions estimated within these segmentation models characterize the respective importance of ontogeny and environmental constraints in tree development at the population scale and have very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. These characteristics may change over tree life, reflecting changes in tree competition. Growth phase duration distributions summarize the joint trajectory of tree ontogeny and environment without requiring tree growth follow-up data for their estimation.

- *Analyzing fruit tree phenology in various climatic conditions from the shoot to the orchard scale* Yann Guédon, Juliano Dutra Schmitz [Universidade Federal de Pelotas, Brazil], Pierre-Eric Lauri [AFEF team, AGAP], Jean-Michel Legave [AFEF team, AGAP], Gustavo Malagui [Universidade Tecnológica Federal do Paraná]

  A first study was conducted at the shoot scale on four apple cultivars characterized by various chilling requirements and grown in two contrasting winter temperature conditions [20]. Our hypothesis was that shoot architecture is strongly affected by winter temperatures determining both the position and budburst of vegetative laterals with a lower effect on their outgrowth. A two-step approach was designed to quantify at the shoot scale i) the branching pattern and ii) two phenological stages of vegetative laterals, budburst and outgrowth. The branching pattern analysis combined various methods: branching sequence alignement and clustering, hidden semi-Markov chain and multiple change point model for segmenting branching sequences into homogeneous zones. A categorical variable, the branching zone, was built to summarize the lateral position along the shoot. It was integrated into the phenological analysis, based on a zero-inflated Poisson model, as a factor together with the cultivar and the winter temperature. We showed in this way that temperature had a main effect on the distribution of vegetative laterals along the shoot. It also strongly affected budburst, which was also affected by the cultivar and the branching zone. The outgrowth of the lateral was not significantly affected by temperature but was significantly affected by the cultivar and the branching zone. Furthermore, the delayed senescence and subsequent leaf persistence during winter, characterizing the apple tree in the mild winter temperature condition, had only a weak effect on the distribution of vegetative laterals and on budburst and lateral outgrowth. The actual shoot architecture and budburst result thus from an ordered sequence of events with a pivotal role of winter temperatures on the dormancy completion of individual lateral buds.

  A second study was conducted at the orchard scale. The time-course variation of dates of flowering stages was established for seventeen chronological sequences corresponding to various apple tree cropping areas in Europe (Belgium, France, Germany, Italy, Switzerland), north Africa (Morocco) and southern Brazil. Our aim was to characterize the relationship between flowering advances in fruit trees and global warming and to compare the northern and the southern hemisphere situations. We applied piecewise constant and linear homoscedastic models to these phenological series. The sudden advance of flowering dates detected at the end of the 1980s in the European locations can be explained by changes in rates for completion of heat requirements, essential to the development of floral primordia within buds. No effect of the global warming could be detected in the Brazilian flowering series and we only found a direct effect of the chilling temperature on the flowering date the same year (the colder the Austral winter, the earlier the flowering date).

- *Investigating how architectural development interfer with epidemics and epidemic control* (Christian Fournier, Corinne Robert [EGC], Guillaume Garin [ITK], Bruno Andrieu [EGC], Christophe Pradal)

  Sustainable agriculture requires the identification of new, environmentally responsible strategies of crop protection. Modelling of pathosystems can allow a better understanding of the major interactions inside these dynamic systems and may lead to innovative protection strategies. In particular, functional–structural plant models (FSPMs) have been identified as a means to optimize the use of architecture-related traits. A current limitation lies in the inherent complexity of this type of modelling, and thus the purpose of this paper is to provide a framework to both extend and simplify the modelling of pathosystems using FSPMs.

  Different entities and interactions occurring in pathosystems were formalized in a conceptual model [21]. A framework based on these concepts was then implemented within the open-source OpenAlea modelling platform, using the platform's general strategy of modelling plant–environment interactions and extending it to handle plant interactions with pathogens. New developments include a generic data structure for representing lesions and dispersal units, and a series of generic protocols to communicate with objects representing the canopy and its microenvironment in the OpenAlea platform. Another development is the addition of a library of elementary models involved in pathosystem modelling. Several plant and physical models are already available in OpenAlea and can be combined in models of pathosystems using this framework approach.

  Two contrasting pathosystems are implemented using the framework and illustrate its generic utility. Simulations demonstrate the framework's ability to simulate multiscaled interactions within pathosystems, and also show that models are modular components within the framework and can be extended. This is illustrated by testing the impact of canopy architectural traits on fungal dispersal. This study provides a framework for modelling a large number of pathosystems using FSPMs. This structure can accommodate both previously developed models for individual aspects of pathosystems and new ones. Complex models are dissassembled into separate *knowledge sources* originating from different specialist areas of expertise and these can be shared and reassembled into multidisciplinary models. The framework thus provides a beneficial tool for a potential diverse and dynamic research community.

5.3. Meristem functioning and development

In axis 2 work focuses on the creation of a *virtual meristem*, at cell resolution, able to integrate the recent results in developmental biology and to simulate the feedback loops between physiology and growth. The approach is subdivided into several sub-areas of research.

5.3.1. Data acquisition and design of meristem models

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging

  Participants: Léo Guignard, Christophe Godin, Christophe Pradal, Grégoire Malandain [Morpheme, Inria], Gaël Michelin [Morpheme, IPL Morphogenetics, Inria], Guillaume Baty, Sophie Ribes [IBC, UM2], Jan Traas [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Yassin Refahi [RDP, ENS].

This research theme is supported by a PhD FRM grant, Jan Traas's ERC, Inria ADT programme and the Morphogenetics Inria Project Lab.

The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline automatically performs a segmentation at cell resolution from 3D or 2D voxel images where the membranes/walls are marked (by a die for example) and makes it possible to follow the lineage of these cells through time [5]. This year, a new version of this pipeline has been developed that uses informations redundancy across the movies and biological knowledge on the segmented organism to constrain and therefore improve the segmentation and the tracking. To test the new pipeline, we used different acquisition protocols and different organisms (floral and apical meristems and the early stages of development of a marine animal *Phallusia mammillata*). The segmentation is corrected a posteriori to deal with imaging artifacts due to uncertainties of acquisition. The image data set on which we develop the methods consists of :

- *Arabidopsis thaliana* shoot apical meristem and primordia with around 6000 cells. The organ is captured from single angle every 4 hours during 2 or 3 days with a confocal microscope (Collaboration Sainsbury lab, Cambridge)
- *Arabidopsis thaliana* flower meristems with around 2000 cells. The organ is also captured from single angle with a confocal microscope (Collaboration RDP Lyon and Sainsbury lab)
- *Phallusia mammillata* embryos with from 32 cells to around 1000 cells. The organism is captured from four different angles every minute during 10 hours with a SPIM (Single Plane Illumination Microscope) (Collaboration CRBM Montpellier / EMBL Heidelberg). This work is developed in the context of the PhD work of Léo Guignard.

To our knowledgeIt is the first time that such high-resolution 4D digital tissues have been generated taking into account the cell shapes, opening the way to quantitative analysis of morphogenesis and tissue deformation at cell resolution.

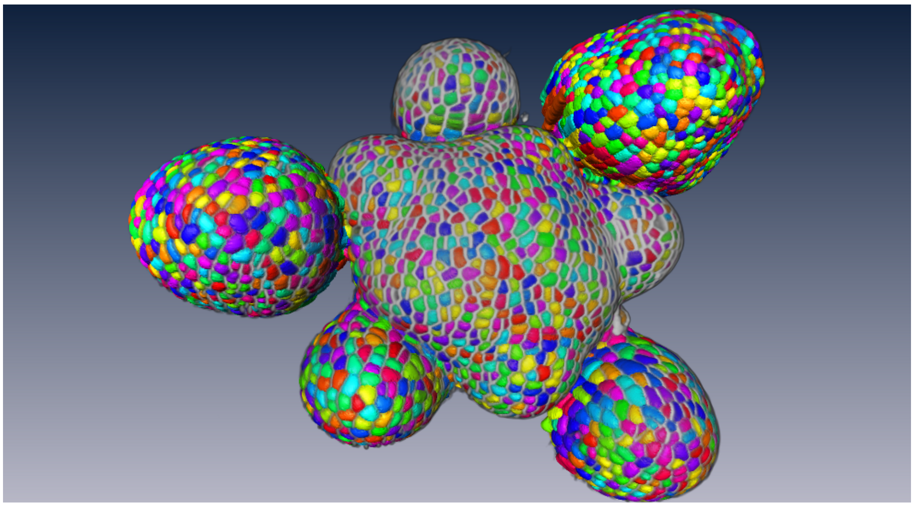


Figure 3. Superimposition of an automatic cell segmentation of an arabidopsis flower meristem using the new MARS pipeline with the original confocal image stack where the membranes are marked.

- *Creating mesh representation of cellular structures* (Guillaume Cerutti, Sohie Ribes, Christophe Godin)

Participants: Guillaume Cerutti, Sophie Ribes, Christophe Godin, Géraldine Brunoud [RDP, ENS], Carlos Galvan-Ampudia [RDP, ENS], Teva Vernoux [RDP, ENS], Yassin Refahi [RDP, ENS, Sainsbury Lab].

This research theme is supported the HFSP project Biosensors.

To produce a more efficient data structure accounting for the geometry of cellular tissues, we studied the problem of reconstructing a mesh representation of cells in a complex, multi-layered tissue structure, based either on membrane/wall images segmented using MARS or on nuclei images of shoot apical meristems. The construction of such mesh structures for plant tissues is currently a missing step in the existing image analysis pipelines. We developed a set of tools to build a triangular mesh surface representing the tissue in 3D, to evaluate the quality of the tissue reconstruction over objective aspects, to optimize a low-quality mesh simultaneously along several criteria, and to go towards a higher-scale representation pulling away from the cell resolution [31]. These methods are used in particular on nuclei images of shoot apical meristems of *Arabidopsis thaliana* to project hormonal information at cell-level on a continuous 3D tissue geometry. This work is carried out in the context of the post-doc of Guillaume Cerutti within the HFSP project BioSensors (Collaboration RDP Lyon).

These tools can produce light discrete representations of the cell interfaces that enables fast visualization, information projection, and quantitative analysis of the tissue, and give way to *in silico* physical and mechanical simulations on real-world data.

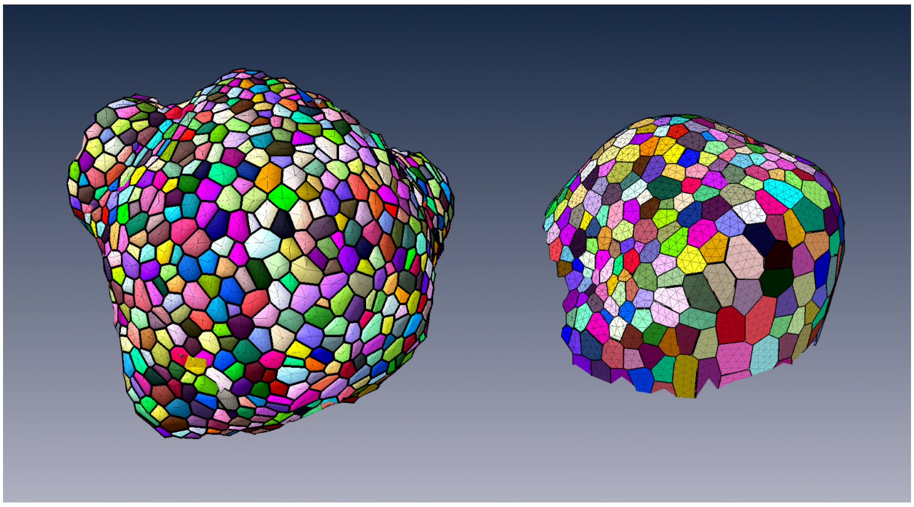


Figure 4. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

- *Design of 3D digital atlases of tissue development*
  Participants: Sophie Ribes, Yassin Refahi [RDP, ENS, Sainsbury Lab], Guillaume Cerutti, Christophe Godin, Christophe Pradal, Christophe Pradal, Frédéric Boudon, Gregoire Malandain [RDP, ENS], Gaël Michelin [RDP, ENS], Guillaume Baty, Jan Traas [RDP, ENS], Teva Vernoux [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Françoise Monéger [RDP, ENS].

  This research theme is supported the Inria Project Lab Morphogenetics, the ADT Mars-Alt and the HFSP project Biosensors.

  To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database [56]. This database makes it possible to store all the collected information on a virtual 3D structure representing a typical organ. Each piece of information has to be located spatially and temporally in the database. Tools to visually retrieve and manipulate the information, quantitatively through space and time are being developed. For this, the 3D structure of a typical organ has been created at the different stages of development of the flower bud. This virtual structure contains spatial and temporal information on mean cell numbers, cell size, cell lineages, possible cell polarization (transporters, microtubules), and gene expression patterns. Such 3D digital atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the digital atlas according to main index keys, in particular spatial and temporal keys. Both a dedicated language and a 3D user interface are being designed to investigate and query the 3D virtual atlas. Current developments of this tool consist in using directly the segmented images produced from laser microscopy to build the atlas. To better represent the development of a biological population, a method to compute an "average" structure is investigated.

5.3.2. Shape analysis of meristems

Participants: Jonathan Legrand, Pierre Fernique, Frédéric Boudon, Yann Guédon, Christophe Godin, Pradeep Das [RDP, ENS], Arezki Boudaoud [RDP, ENS].

At cellular resolution, we studied the organization of cells in the meristems. The MARS-ALT pipeline provides rich spatio-temporal data sets for analyzing the development of meristems. A first step consisted of designing a dedicated graph for efficiently representing the spatial (adjacency between cells) and temporal (cell division) relationships between cells. Various variables can be attached either to the vertices (e.g. cell volume, inertia axes) or the edges (e.g. wall surface, distance between cell centroids). This graph may be augmented by new variables resulting from various spatial or temporal filtering (e.g. cell volumetric growth). Looking at homogeneous regions in the variable value space, cellular patterns can be identified. This work was developed in the context of the PhD of Jonathan Legrand with contributions of Pierre Fernique, another PhD student, that has been defended this year.

Considering the highly-structured nature of our data (time and space structuring) and the potential diversity and heterogeneity of possible cell descriptors, we developed two complementary approaches:

- A first one that favours the spatial structuring: In this approach, the cell neighbourhood and the cell descriptors are jointly taken into account in a clustering approach whose objective is to identify a small number of clusters corresponding to well-defined cell identities. Once the cells have been labelled using the clustering algorithm, cell generation distributions are estimated on the basis of the labelled lineage trees.
- A second one that favours the temporal structuring: In this approach, the data of interest are lineage forest and the only spatial structuring taken into account corresponds to siblings with respect to a given parent cell. In a first step, cell identities are inferred on the basis of the cell descriptors taking into account lineage relationships using hidden Markov tree models and the spatial regions that emerge from the cell identity labelling are then characterized. This second approach is supported by the fact that cell topology is only affected by division which makes highly relevant the local spatial information taken into account in this approach.

5.3.3. Mechanical model

Participants: Jean-Philippe Bernard, Olivier Ali, Christophe Godin, Benjamin Gilles, Frédéric Boudon, Ibrahim Cheddadi, Jan Traas [ENS-Lyon], Olivier Hamant [ENS-Lyon], Arezki Boudaoud [ENS-Lyon].

This research theme is supported by the Inria Project Lab Morphogenetics and the Jan Traas's ERC.

The rigid cell walls that surround plant cells are responsible for the acquisition of organ shapes. These walls are submitted to stresses due to cell turgor pressure. Wall deformation is caused by the turgor forces in the cell walls. Wall synthesis is triggered by these wall deformations when some specific threshold is exceeded. The final shape of the tissue integrates mechanically all the local deformations of each cell.

To quantify this growth process at the level of a multicellular tissue, we developed a model of growth that integrates mechanical forces development at cellular resolution. In this model, walls are characterized by their mechanical properties, e.g. elasticity, extensibility and anisotropy. For this, we used a tensorial approach to describe both tissue deformation and stresses. Deformations were decomposed into elementary transformations that can be related to underlying biological processes. However, we showed that the observed deformations does not map directly local growth instructions given by genes and physiology in each cell. Instead, the growth is a two-stage process where genes are specifying how cell walls should yield to mechanical stresses. In this way, different regions in the tissue with different cell identities can have different growth properties. The final shape of the tissue results from the integration of all these mechanical properties and stresses at organ level under the growth force due to turgor pressure at tissue scale.

A paper describing the mechanical model and its application to model primorium formation in the shoot apical meristem will appear in PLoS Computational Biology in 2015. We used this framework to investigate the influence of a specific signalling cascade (the ABP1- Kat1 signalling pathway) and its putative mechanical consequences on primordium initiation [25]. A review of the different mechanical concepts underlying plant morphogenesis has been carried out in [12].

In our first approach, the mechanical model rely on a finite element method (FEM) to describe the deformation of the tissue. In FEM, the tissue is represented by a mesh. The positions of the vertices at each time step are estimated from a linear system. If the tissue is big or if the mesh is fine, the linear system can be large and thus leads to computational overheads. An alternative way to classical FEM is to use a meshless method where the deformation of the tissue can be characterized by a linear combination of deformations of a finite and small set of frames. Because shape functions are no longer defined on each element but on the whole tissue, they have to be updated at each growth step by estimating a new rest configuration. With meshless method, the discretization of the system can be dynamically updated parsimoniously according to the precision required to model the emergence of shapes (PhD work of Jean-Philippe Bernard).

5.3.4. Gene regulatory networks: Design of a genetic model of inflorescence development.

Participants: Eugenio Azpeitia, Christophe Godin, François Parcy, Etienne Farcot.

This research theme is supported by the Inria Project Lab Morphogenetics.

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression.

We studied the regulatory network that controls the flowering transition during morphogenesis. To overcome the network complexity and integrate this regulation during ontogenesis, we have developed a first model of the control of floral initiation by genes, and in particular the situation of cauliflower mutants, in which the repeatedly meristem fails in making a complete transition to the flower. Three different network models were done and validate. A first Boolean version, a second fuzzy logic and an ODEs models were studied. The models are able to correctly recover the gene steady states observed in the meristems during the flower transitions, the gene transitions and the mutant effects. Importantly, the model is able to explain the cauliflower mutants. This work couples models at different scales, since the gene regulatory network is used as a decision module in an L-system model of the inflorescence architecture. This mixed model has led us to make different hypotheses about gene interactions and hormonal regulation. First predictions about gene actors controling the passage to flower could be verified. Finally, some links between gene regulation and plant growth have been identified. These links can be experimentally tested which could lead to a first integrated picture of flower development could be reached in the context of Eugenio Azpeitia postdoc.

5.3.5. Model integration

Participants: Frédéric Boudon, Christophe Godin, Guillaume Baty, Jan Traas.

This research theme is supported by the Morphogenetics Inria Project Lab.

Our approach consists of building a programmable tissue which is able to accept different modeling components. This includes a central data structure representing the tissue in either 2-D or 3-D, which is able to grow in time, models of gene activity and regulation, models of signal exchange (physical and chemical) between cells and models of cell cycle (which includes cell division). For each modeling component, one or several approaches are investigated in depth, possibly at different temporal and spatial scales, using the data available from the partners (imaging, gene networks, and expression patterns). Approaches are compared and assessed on the same data. The objective of each submodel component will be to provide plugin components, corresponding to simplified versions of their models if necessary, that can be injected in the programmable tissue platform. This work is developed in collaboration with the RDP group at ENS-Lyon [58] and the CPIB group in Nottingham, UK [47].

One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development. This platform will be used to carry out integration of the different models developed in this research axis. The platform is based on *OpenAlea*. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform. Currently, robust geometric operations such as division are implemented and tested. Moreover, a redesign of the structure based on more elaborated formalisms such as combinatorial maps is investigated.

5.4. Multi-scale models and analysis: from cells to plant architecture (and back)

5.4.1. Modeling water transport in roots

Participants: Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, Christophe Maurel [BPMP].

This research theme is supported by the ANR project HydroRoot.

A model of Arabidopsis thaliana root hydraulics at the cellular level was developped in the OpenAlea modeling platform. The model relies on the integration throughout root architecture of elementary hydraulic components. Each component integrates local radial and axial water flows. Axial hydraulic conductivity is calculated according to Poiseuille's law, based on local size of xylem vessels. Radial hydraulic conductivity is determined in part by aquaporin activity and was set constant throughout root architecture in the first model versions. In its current state, the model is parameterized using architectural, tissular and physiological data that were experimentally determined in the Aquaporin group at BPMP. The architectural reconstruction of the root system is based on a tridimensional multi-scale tree graph (MTG). The current model is capable of predicting the water flow that is transported by a root system in the standard experimental conditions used in the Aquaporin group. This model was used to perform sensitivity analyses and determine the respective contributions to root hydraulic dynamics of various biological parameters (axial and radial hydraulic conductivites, root architecture). One major finding is that the root hydraulic conductivity (Lpr) computed from the model is highly dependent on root architecture. This is due to the limiting role of axial (xylem) conductance, one feature that had been neglected in previous representations of root water transport. The radial hydraulic conductivity may primarily be limiting in conditions of Lpr inhibition, since its increase from values in control roots has marginal effects on Lpr. A new set of experimental data including root diameter repartitions in wild-type plants, and xylem vessel diameters in mutants with altered xylem morphology (irx3, esk1) will be used to implement the model. Root cell hydraulic conductivities will also be measured in these and aquaporin mutant phenotypes. Our aim is to check whether, based on anatomical and morphological data, the model can properly predict the radial hydraulic conductivity of these genotypes.

5.4.2. Mechanical modeling of fruit growth

Participants: Ibrahim Cheddadi [Inra, Avignon], Mik Cieslak [U. Calgary], Frédéric Boudon, Valentina Baldazzi [Inra, Avignon], Nadia Bertin [Inra, Avignon], Michel Genard [Inra, Avignon], Christophe Godin.

This research theme is supported by the Agropolis project MecaFruit3D.

Understanding the controlling factors of fruit quality development is challenging, because fruit quality results from the interplay between physical and physiological processes that are under the control of genes and the environment. Although process-based models have been used to make significant progress in understanding these factors, they ignored to a large extent the shape and internal structure of the fruit, as well as mechanical interactions between tissue parts that are essential to properly model growth.

To help characterizing the effects of fruit shape and internal structure on quality, the creation of a 3D virtual fruit model that integrates fruit structure and function with growth governed by environmental inputs is being investigated, combining two tools previously developed in the team: on the one hand, a modeling pipeline that creates a 3D volumetric mesh of the internal fruit structure, including vasculature (see section 3), and couples it with water and carbon transport; on the other hand, a mechanical description of the growth of plant tissues (see section 5.3.3): growth is related to the extension of the cell walls, which is triggered when the so-called turgor pressure inside the cells exceeds a given threshold. The global shape of the tissue integrates mechanically all the local deformations of each cell.

In order to couple these two aspects of plant growth, we describe how volume variations are constrained by fluxes of matter, and how these fluxes depend on mechanical and physiological parameters. The corresponding set of equations are resolved thanks to the SOFA finite elements software.

This approach will be applied to study tomato fruit. Once the model is calibrated and evaluated, our approach will be suitable for studying the effects of internal fruit heterogeneity and overall shape on fruit quality development.

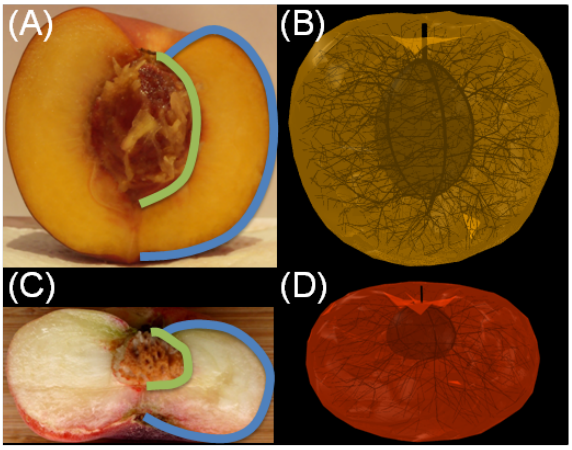


Figure 5. Virtual models of peaches reconstructed from images with simulated vasculatur to simulate carbon and water transport in the fruit

5.4.3. Analyzing root growth and branching

Participants: Beatriz Moreno Ortega, Sixtine Passot, Yann Guédon, Laurent Laplaze [IRD, DIADE], Mikaël Lucas [IRD, DIADE], Bertrand Muller [INRA, LEPSE].

This research theme is supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. SmartRoot). The analysis of complex root phenotyping data is thus a new challenge in developmental biology.

We aim at developing a pipeline of methods for analyzing root systems at three scales:

1. tissular scale to identify and characterize the meristem, elongation and mature zones along a root using piecewise heteroscedastic linear models,
2. individual root scale to analyze the dynamics of root elongation,
3. root system scale to analyze the branching structure.

This pipeline of analysis methods will be applied to different species (maize, millet and *arabidopsis*) and for different biological objectives (study of genetic diversity for millet and of metabolic and hormonal controls of morphogenesis for maize).

5.4.4. Analyzing shoot and leaf elongation

Participants: Maryline Lièvre, Yann Guédon, Leo Guignard, Christine Granier [INRA, LEPSE].

This research theme is supported by one PhD programme.

This study is based on the observation of a lack of methods enabling the integrated analysis of the processes controlling the vegetative development in *Arabidopsis thaliana*. We developed a pipeline of analysis methods combining image analysis techniques and statistical models to integrate the measurements made at the leaf and shoot scales. Semi-Markov switching models were built for different genotypes, allowing a more thorough characterization of the studied mutants. These models validated the hypothesis that the rosette can be structured into successive developmental phases that could change depending on the genotype. They also highlighted the structuring role of the abaxial trichome trait, although the developmental phases cannot be explained entirely by this trait. We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images based on the ilastik software, and the analysis of the obtained cell areas using a gamma mixture model whose parameters of gamma components are tied by a scaling rule. This model allowed us to estimate the distribution of the number of endocycles. We highlighted in this way that the mean number of endocycles changes drastically with leaf rank. Finally, we built a multi-scale model that integrates tissular, morphological, dynamical and dimension traits for each successive leaf along the shoot. This model gave us for the first time an integrative view of the development of the *Arabidopsis* rosette.

5.4.5. Analyzing perturbations in Arabidopsis thaliana phyllotaxis

Participants: Yassin Refahi, Fabrice Besnard, Yann Guédon, Christophe Godin, Etienne Farcot, Teva Vernoux [RDP, ENS].

This research theme has been supported by iSam, IBC and the Inria Project Lab Morphogenetics.

The geometric arrangement of lateral organs along plant stems, named phyllotaxis, shows a variety of striking patterns with remarkable regularities and symmetries. This has interested biologists, physicists, mathematicians and computer scientists for decades. These studies have lead to a commonly accepted standard interpretation of phyllotaxis that postulates that organs inhibit the formation of new organs in their vicinity. At a molecular scale, these inhibitory fields have been shown to result from the spatio-temporal distribution of the plant hormone auxin. This model theoretically explains a large part of the diversity of phyllotactic patterns observed in plants.

The cytokinin hormones are known to play a significant role in the regulation of phyllotaxis. Fabrice Besnard and Teva Vernoux realized that *Arabidopsis thalianaahp6* mutants, which are perturbed in the cytokinin signaling pathway, showed unusual chaotic perturbations of the phyllotaxis at macroscopic level.

In order to characterize these perturbations, we designed a pipeline of models and methods [60], [53] which relies of combinatorial and statistical techniques. Using this pipeline of methods, we have shown that the perturbation patterns in both wild-type and mutant plants can be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs.The number of successive synchronized organs between two permutations reveals unexpected patterns that depend on the nature of the preceding permutation (2- or 3-permutation). We identified significant individual deviations of the level of baseline segments with reference to 137.5°, which confirms theoretical model predictions. Finally, we highlighted a marked relationship between permutation of organs and defects in the elongation of the internodes in between these organs.

We then looked at the origin of these permutations using confocal microscopy and realized that organs were in fact frequently co-initiated in the mutant, leading after development randomly in half of the cases to permutations. We concluded that the mutant is actually perturbed in the time between consecutive organ initiation (i.e. the plastochrone), while relative angular positions are not affected. After closer inspection, we realized that the mutated gene encodes a protein diffusing from the organs and creating a field around the organs that regulates the plastochrone. We could demonstrate that in the mutant, the absence of this field leads to co-initiations and subsequently to the observed permutations.

Altogether, this study sheds a new light on our interpretation of phyllotaxis, revisiting the standard model and suggesting that several fields based on auxin and cytokinin with different properties are required to provide robustness to phyllotaxis. An overview of this work has been published in the journal Nature [15].

5.4.6. A stochastic model of phyllotaxis

Participants: Yassin Refahi, Christophe Godin, Etienne Farcot, Teva Vernoux [RDP, ENS].

This research theme has been supported by IBC and the Inria Project Lab Morphogenetics.

To proceed further and find a mechanistic interpretation of these results, we are currently developing a stochastic extension of the standard model of phyllotaxis. We first analyzed the properties of the inhibitory fields created by the existing primordia on the initiation of new promordia, and concluded that the angular positions of organs are very robust to perturbations while plastochrons may be dramatically affected. This suggested that there exists a strong decoupling between space and time in the patterning process. To account for this observation, we modeled the perception of the initiation signal by cells using stochastic processes coupled with the intensity of inhibitory fields and showed that the observed permutation patterns emerge spontaneously from this purely local processes. A paper describing this model will be submitted for publication in 2015.

5.4.7. The role of auxin and sugar in rose bud outgrowth control

Participants: Jessica Bertheloot [INRA, Angers], Frédéric Boudon, Christophe Godin.

Auxin in the stem is known to be a key regulator of apical dominance. Over the last decades, many studies have been undertaken to understand its action mode, which is indirect because auxin in the main stem does not enter into the bud. Recently, apical dominance over basal buds in pea has been related to low sugar availability caused by high sugar demand of growing apical organs. Auxin and sugar are two signals regulating the entrance of bud into sustained growth in opposite ways. Recently it has been demonstrated that sugar effect on bud outgrowth was preceded by a modification of the hormonal network involved in auxin effect, which suggests that auxin and sugar pathways do interact in a non-trivial way. But auxin and sugar effects have been studied separately until now. In this work, we investigate what is the combined effect of sugar and auxin on bud outgrowth, and how they integrate to control bud entrance into sustained growth. For this, a series of experiments has been carried on a single-node cuttings of Rosa hybrida grown in vitro in which different combinations of sugar and auxin levels have been tested. A model the regulatory networks controling stem-bud molecular interaction is currently being investigated.

ARAMIS Project-Team

6. New Results

6.1. Highlights of the Year

ARAMIS has contributed to the special issue on "Complex network theory and the brain" in the prestigious journal of Philosopical Transactions of the Royal Society, Series B. This work was featured by the ICM (http://icm-institute.org/en/news/complex-network-theory-and-the-brain?lang=en) and Inria (http://www.inria.fr/en/centre/paris-rocquencourt/news/complex-network-theory-and-the-brain).

6.2. Detection of volume loss in hippocampal layers in Alzheimer's disease using 7 T MRI

Participants: Claire Boutet, Marie Chupin, Stéphane Lehéricy, Linda Marrakchi-Kacem, Stéphane Epelbaum, Cyril Poupon, Christopher Wiggins, Alexandre Vignaud, Dominique Hasboun, Bénédicte Desfontaines, Olivier Hanon, Bruno Dubois, Marie Sarazin, Lucie Hertz-Pannier, Olivier Colliot [Correspondant].

In Alzheimer's disease (AD), the hippocampus is an early site of tau pathology and neurodegeneration. Histological studies have shown that lesions are not uniformly distributed within the hippocampus. Moreover, alterations of different hippocampal layers may reflect distinct pathological processes. 7 T MRI dramatically improves the visualization of hippocampal subregions and layers. In this study, we aimed to assess whether 7 T MRI can detect volumetric changes in hippocampal layers in vivo in patients with AD. We studied four AD patients and seven control subjects. MR images were acquired using a whole-body 7 T scanner with an eight channel transmit-receive coil. Hippocampal subregions were manually segmented from coronal T2*-weighted gradient echo images with $0.3 \times 0.3 \times 1.2$ mm3 resolution using a protocol that distinguishes between layers richer or poorer in neuronal bodies (Figure 1). Five subregions were segmented in the region of the hippocampal body: alveus, strata radiatum, lacunosum and moleculare (SRLM) of the cornu Ammonis (CA), hilum, stratum pyramidale of CA and stratum pyramidale of the subiculum. We found strong bilateral reductions in the SRLM of the cornu Ammonis and in the stratum pyramidale of the subiculum ($p < 0.05$), with average cross-sectional area reductions ranging from -29% to -49%. These results show that it is possible to detect volume loss in distinct hippocampal layers using segmentation of 7 T MRI. 7 T MRI-based segmentation is a promising tool for AD research.

More details in [3].

6.3. White matter lesions in patients with frontotemporal lobar degeneration due to progranulin mutations

Participants: Paola Caroppo, Isabelle Le Ber, Agnès Camuzat, Fabienne Clot, Lionel Naccache, Foudil Lamari, Anne Bertrand, Serge Belliard, Olivier Colliot [Correspondant], Alexis Brice.

Mutations in the progranulin (GRN) gene are responsible for 20% of familial cases of frontotemporal dementias. All cause haploinsufficiency of progranulin, a protein involved in inflammation, tissue repair, and cancer. Carriers of the GRN mutation are characterized by a variable degree of asymmetric brain atrophy, predominantly in the frontal, temporal, and parietal lobes. We described four GRN mutation carriers with remarkable widespread white matter lesions (WML) associated with lobar atrophy shown on magnetic resonance imaging. The WML were predominantly in the frontal and parietal lobes and were mostly confluent, affecting the periventricular subcortical white matter and U-fibers. In all patients, common vascular, metabolic, inflammatory, dysimmune, and mitochondrial disorders were excluded and none had severe vascular risk factors. Our data suggest that white matter involvement may be linked to progranulin pathological processes in a subset of GRN mutation carriers. The plasma progranulin measurement, which is predictive of GRN mutations, and GRN sequencing should thus be included in investigations of patients with frontotemporal lobar degenerations who show unusual white matter hyperintensities and atrophy on magnetic resonance imaging.

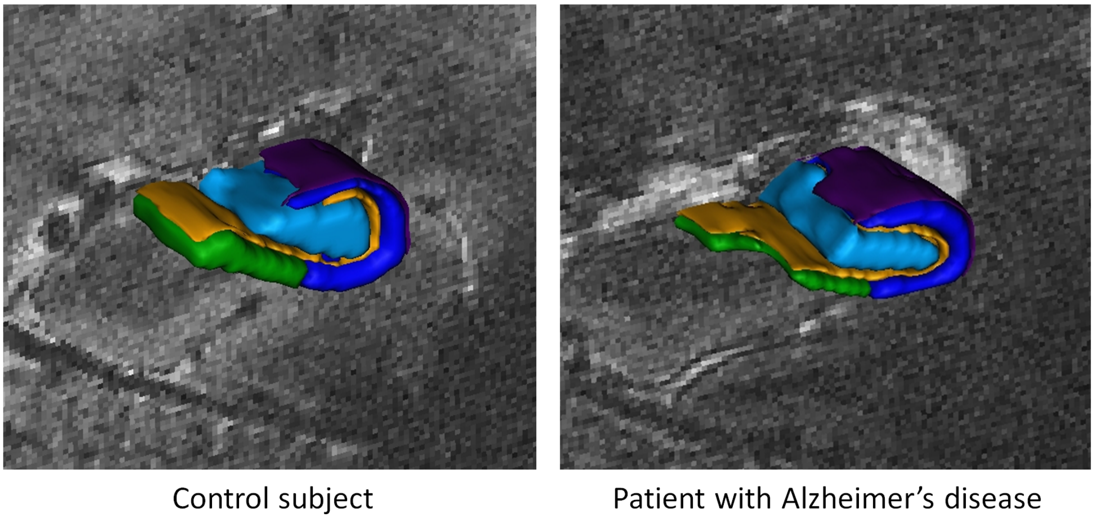


Figure 1. Segmentation of hippocampal layers using in vivo 7 Tesla MRI. were performed on the second echo image. Left panel: control subject. Right panel: patient with Alzheimer’s disease. Purple, alveus; dark blue, stratum pyramidale of CA1-3; yellow, strata radiatum,lacunosum and moleculare of CA1-3, strata lacunosum and moleculare of the subiculum and stratum moleculare of gyrus dentatus; cyan, stratum pyramidale of CA4 and stratum granulosum and polymorphic layer of gyrus dentatus; green, stratum pyramidale of the subiculum.

More details in [4].

6.4. Template-based morphometry using diffeomorphic iterative centroids

Participants: Claire Cury [Correspondant], Joan Glaunès, Marie Chupin, Olivier Colliot.

A common approach for the analysis of anatomical variability relies on the estimation of a representative template of the population, followed by the study of this population based on the parameters of the deformations going from the template to the population. The Large Deformation Diffeomorphic Metric Mapping framework is widely used for shape analysis of anatomical structures, but computing a template with such framework is computationally expensive. We proposed a fast approach for template-based analysis of anatomical variability. The template is estimated using an iterative approach which quickly provides a centroid of the population. Statistical analysis is then performed using principal component analysis on the initial momenta that define the deformations between the centroid and each subject of the population. This approach was applied to the analysis of hippocampal shape on 80 patients with Alzheimer's Disease and 138 controls from the ADNI database.

More details in [22] and [36].

6.5. Structural connectivity differences in left and right temporal lobe epilepsy

Participants: Pierre Besson, Vera Dinkelacker [Correspondant], Romain Valabrègue, Lionel Thivard, Xavier Leclerc, Michel Baulac, Daniela Sammler, Olivier Colliot, Stéphane Lehéricy, Séverine Samson, Sophie Dupont.

Our knowledge on temporal lobe epilepsy (TLE) with hippocampal sclerosis has evolved towards the view that this syndrome affects widespread brain networks. Diffusion weighted imaging studies have shown alterations of large white matter tracts, most notably in left temporal lobe epilepsy, but the degree of altered connections between cortical and subcortical structures remains to be clarified. We performed a whole brain connectome analysis in 39 patients with refractory temporal lobe epilepsy and unilateral hippocampal sclerosis (20 right and 19 left) and 28 healthy subjects. We performed whole-brain probabilistic fiber tracking using MRtrix and segmented 164 cortical and subcortical structures with Freesurfer. Individual structural connectivity graphs based on these 164 nodes were computed by mapping the mean fractional anisotropy (FA) onto each tract. Connectomes were then compared using two complementary methods: permutation tests for pair-wise connections and Network Based Statistics to probe for differences in large network components. Comparison of pair-wise connections revealed a marked reduction of connectivity between left TLE patients and controls, which was strongly lateralized to the ipsilateral temporal lobe. Specifically, infero-lateral cortex and temporal pole were strongly affected, and so was the perisylvian cortex. In contrast, for right TLE, focal connectivity loss was much less pronounced and restricted to bilateral limbic structures and right temporal cortex. Analysis of large network components revealed furthermore that both left and right hippocampal sclerosis affected diffuse global and interhemispheric connectivity. Thus, left temporal lobe epilepsy was associated with a much more pronounced pattern of reduced FA, that included major landmarks of perisylvian language circuitry. These distinct patterns of connectivity associated with unilateral hippocampal sclerosis show how a focal pathology influences global network architecture, and how left or right-sided lesions may have differential and specific impacts on cerebral connectivity.

More details in [2].

6.6. Morphometry of anatomical shape complexes with dense deformations and sparse parameters

Participants: Stanley Durrleman [Correspondant], Marcel Prastawa, Nicolas Charon, Julie Korenberg, Sarang Joshi, Guido Gerig, Alain Trouvé.

We propose a generic method for the statistical analysis of collections of anatomical shape complexes, namely sets of surfaces that were previously segmented and labeled in a group of subjects. The method estimates an anatomical model, the template complex, that is representative of the population under study. Its shape reflects anatomical invariants within the dataset. In addition, the method automatically places control points near the most variable parts of the template complex. Vectors attached to these points are parameters of deformations of the ambient 3D space. These deformations warp the template to each subject's complex in a way that preserves the organization of the anatomical structures. Multivariate statistical analysis is applied to these deformation parameters to test for group differences. Results of the statistical analysis are then expressed in terms of deformation patterns of the template complex, and can be visualized and interpreted. The user needs only to specify the topology of the template complex and the number of control points. The method then automatically estimates the shape of the template complex, the optimal position of control points and deformation parameters. The proposed approach is completely generic with respect to any type of application and well adapted to efficient use in clinical studies, in that it does not require point correspondence across surfaces and is robust to mesh imperfections such as holes, spikes, inconsistent orientation or irregular meshing.

The approach is illustrated with a neuroimaging study of Down syndrome (DS). Results demonstrate that the complex of deep brain structures shows a statistically significant shape difference between control and DS subjects. The deformation-based modeling is able to classify subjects with very high specificity and sensitivity, thus showing important generalization capability even given a low sample size. We show that results remain significant even if the number of control points, and hence the dimension of variables in the statistical model, are drastically reduced. The analysis may even suggest that parsimonious models have an increased statistical performance.

The method has been implemented in the software Deformetrica, which is publicly available at www.deformetrica.org

More details in [14].

6.7. Iconic-Geometric Nonlinear Registration of a Basal Ganglia Atlas for Deep Brain Stimulation Planning

Participants: Ana Fouquier, Stanley Durrleman, Jérôme Yelnik, Sara Fernandez-Vidal, Eric Bardinet.

We evaluated a nonlinear registration method for warping a 3D histological atlas of the basal ganglia into patient data for deep brain stimulation (DBS) planning. The power of the method is the possibility to combine iconic registration with geometric constraints under a unified diffeomorphic framework. This combination aims to ensure robust and accurate atlas-to-patient warping and anatomy-preserving deformations of stimulation target nuclei. A comparison of the method with a state-of-the-art diffeomorphic registration algorithm reveals how each approach deforms low-contrasted image regions where DBS target nuclei often lie. The technique is applied to T1-weighted magnetic resonance images from a cohort of Parkinsonian subjects, including subjects with standard-size and large ventricles. Results illustrate the effects of iconic or geometric registration alone, as well as how both constraints can be integrated in order to contribute for registration precision enhancement. See Fig. 2 .

More details in [25].

6.8. Evaluation of morphometric descriptors of deep brain structures for the automatic classification of patients with Alzheimer's disease, mild cognitive impairment and elderly controls

Participants: Alexandre Routier [correspondant], Pietro Gori, Ana Fouquier, Sophie Lecomte, Olivier Colliot, Stanley Durrleman.

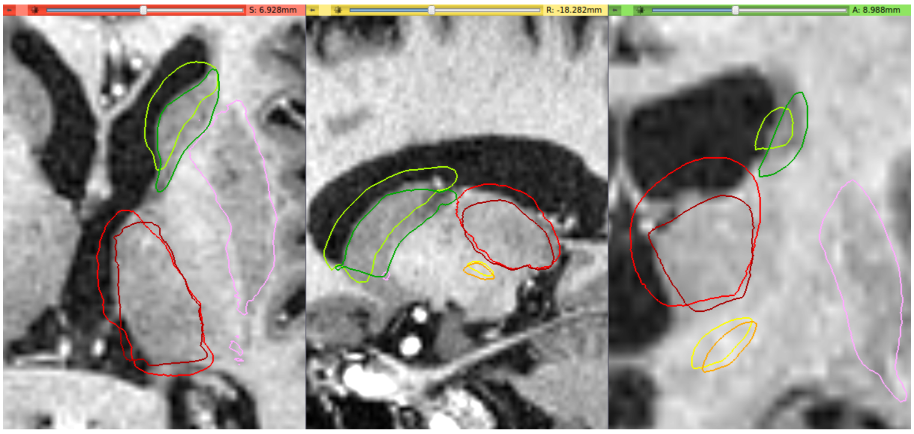


Figure 2. Superimposition of deformed meshes of the histological atlas with a patient pre-operative MRI. Meshes in bright colors result from a block-matching algorithm based on image intensity. Meshes in dark colors result from our iconic-geometric approach with non-linear deformation. We observe a better alignment of the structures, as well as a realistic deformation of the sub-thalamic nucleus (in yellow/orange), which is not visible in the image and therefore has not been taken into account for estimating the optimal deformation. This nucleus is the stimulation target for patients with Parkinson's disease.

We participated in the Computer-Aided Diagnosis of Dementia based on structural MRI data (http://caddementia.grand-challenge.org/). Our approach was to select shapes of 12 brain structures: the caudate nucleus, putamen, pallidum, thalamus, hippocampus and amygdala of each hemisphere. The structure segmentation was based on a FreeSurfer segmentation and the marching-cubes algorithm was used to get 3D triangular meshes. Using our software Deformetrica, anatomical models (mean shape and typical variations) of these brain structures were built for patients with Alzheimer's disease (AD), Mild Cognitive Impairments (MCI) and cognitively normal controls (CN) based on the data of 509 ADNI subjects. The models for AD, MCI and CN were registered to the test subjects by maximizing the likelihood of the test image to be derived from each model. The final classification was made by thresholding this criterion taking into account the co-variance of the deformation parameters. The thresholds were either optimized on the ADNI data or on the provided training data. The method was fully automatic and the computation time was 4 days for training the anatomical models plus 11 hours per subject for registration and classification. For the 30 training subjects, the algorithm had accuracies of 73% (if optimized on training data) and 50% (if optimized on ADNI data). On the test set of 354 images, our method yields an accuracy of 49.2% (43.5 - 54.2), true positive fraction of 94.6% (89.8 - 97.7) for CN, 11.5% (6.2 - 17.7) for MCI and 36.9% (27.4 - 46.5) for AD.

Our participation to this challenge was the opportunity to test our software Deformetrica for classification tasks. It ran on more than 800 images, thus showing its ability to deal with large data sets.

More details in [27].

6.9. A Prototype Representation to Approximate White Matter Bundles with Weighted Currents

Participants: Pietro Gori [correspondant], Olivier Colliot, Linda Marrakchi-Kacem, Fabrizio de Vico Fallani, Mario Chavez, Sophie Lecomte, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman.

Quantitative and qualitative analysis of white matter fibers resulting from tractography algorithms is made difficult by their huge number. To this end, we propose an approximation scheme which gives as result a more concise but at the same time exhaustive representation of a fiber bundle. It is based on a novel computational model for fibers, called weighted currents, characterized by a metric that considers both the pathway and the anatomical locations of the endpoints of the fibers. Similarity has therefore a twofold connotation: geometrical and related to the connectivity. The core idea is to use this metric for approximating a fiber bundle with a set of weighted prototypes, chosen among the fibers, which represent ensembles of similar fibers. The weights are related to the number of fibers represented by the prototypes. The algorithm is divided into two steps. First, the main modes of the fiber bundle are detected using a modularity based clustering algorithm. Second, a prototype fiber selection process is carried on in each cluster separately. This permits to explain the main patterns of the fiber bundle in a fast and accurate way. See Fig. 3

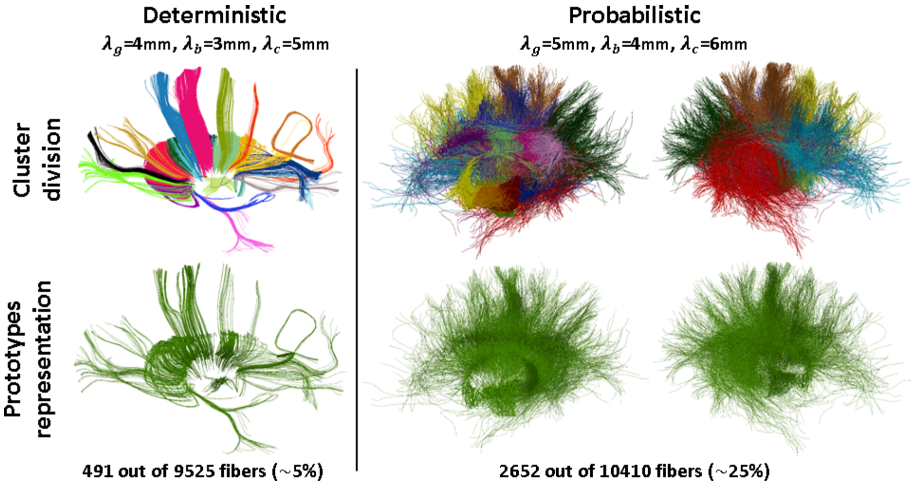


Figure 3. Illustration of our method to cluster fibers and approximate clusters based on a weighted currents metric, which measures differences in the locations of fibers extremities and the geometry of their pathway. 2 examples are shown using fibers from a deterministic tractography (left) and probabilistic tractography (right). Clustering (top row) and approximation of fibers within each cluster (bottom row) are shown.

More details in [24].

6.10. Non-parametric resampling of random walks for spectral network clustering

Participants: Fabrizio de Vico Fallani [correspondant], Vincenzo Nicosia, Vito Latora, Mario Chavez.

Parametric resampling schemes have been recently introduced in complex network analysis with the aim of assessing the statistical significance of graph clustering and the robustness of community partitions. We proposed a method to replicate structural features of complex networks based on the non-parametric resampling of the transition matrix associated with an unbiased random walk on the graph. We tested this bootstrapping technique on synthetic and real-world modular networks and we showed that the ensemble of replicates obtained through resampling can be used to improve the performance of standard spectral algorithms for community detection.

More details in [10].

6.11. Graph analysis of functional brain networks: practical issues in translational neurosciences

Participants: Fabrizio de Vico Fallani [correspondant], Sophie Achard, Jonas Richiardi, Mario Chavez.

The brain can be regarded as a network: a connected system where nodes, or units, represent different specialized regions and links, or connections, represent communication pathways. From a functional perspective, communication is coded by temporal dependence between the activities of different brain areas. In the last decade, the abstract representation of the brain as a graph has allowed to visualize functional brain networks and describe their non-trivial topological properties in a compact and objective way. Nowadays, the use of graph analysis in translational neuroscience has become essential to quantify brain dysfunctions in terms of aberrant reconfiguration of functional brain networks. Despite its evident impact, graph analysis of functional brain networks is not a simple toolbox that can be blindly applied to brain signals. On the one hand, it requires the know-how of all the methodological steps of the pipeline that manipulate the input brain signals and extract the functional network properties. On the other hand, knowledge of the neural phenomenon under study is required to perform physiologically relevant analysis. The aim of our work is to provide practical indications to make sense of brain network analysis and contrast counterproductive attitudes.

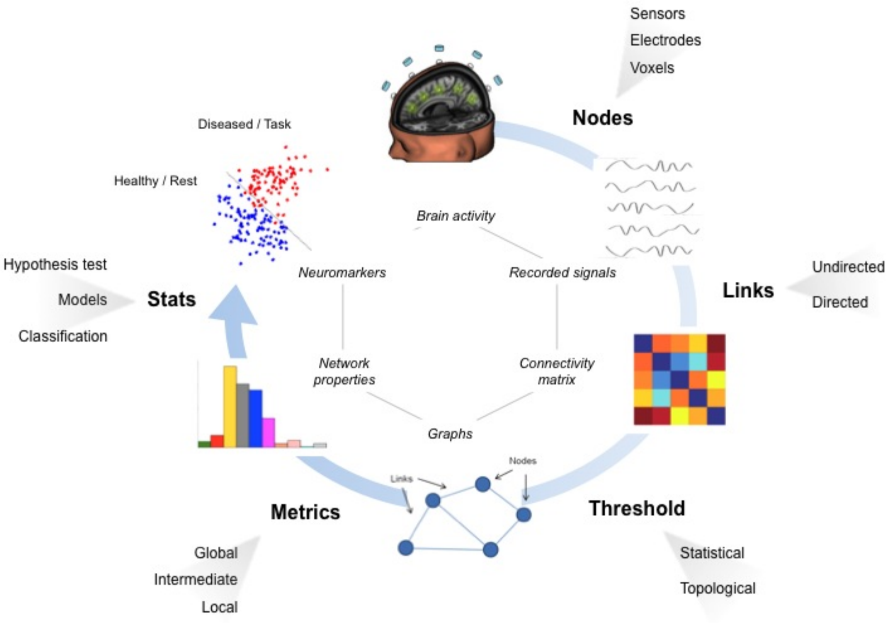


Figure 4. Processing pipeline for functional brain connectivity modeling and analysis. Nodes correspond to specific brain sites according to the used neuroimaging technique. Links are estimated by measuring the functional connectivity (FC) between the activity of brain nodes; this information is contained in a connectivity matrix. By means of filtering procedures, based on thresholds, only the most important links constitute the brain graph. The topology of the brain graph is quantified by different graph metrics (or indices) that can be represented as numbers (e.g. the colored bars). These graph indices can be input to statistical analysis to look for significant differences between populations/conditions (e.g. red points correspond to brain graph indices of diseased patients or tasks, blue points stand for healthy subjects or resting states).

More details in [11].

6.12. Hierarchy of neural organisation in the zebra fish spinal cord: causality analysis of in-vivo calcium imaging data

Participants: Fabrizio de Vico Fallani [correspondant], Martina Corazzol, Jnena Sternberg, Kevin Fidelin, Claire Wyart, Mario Chavez.

The recent development of genetically encoded calcium indicators enables monitoring in vivo the activity of neuronal populations. Most analysis of these calcium transients relies on linear regression analysis based on the sensory stimulus applied or the behavior observed. To estimate the basic properties of the functional neural circuitry, we propose a network-based approach based on calcium imaging recorded at single cell resolution. Differently from previous analysis based on cross-correlation, we used Granger causality estimates to infer activity propagation between the activities of different neurons. The resulting functional networks were then modeled as directed graphs and characterized in terms of connectivity and node centralities. We applied our approach to calcium transients recorded at low frequency (4 Hz) in ventral neurons of the zebrafish spinal cord at the embryonic stage when spontaneous coiling of the tail occurs. Our analysis on population calcium imaging data revealed a strong ipsilateral connectivity and a characteristic hierarchical organization of the network hubs that supported established propagation of activity from rostral to caudal spinal cord. Our method could be used for detecting functional defects in neuronal circuitry during development and pathological conditions.

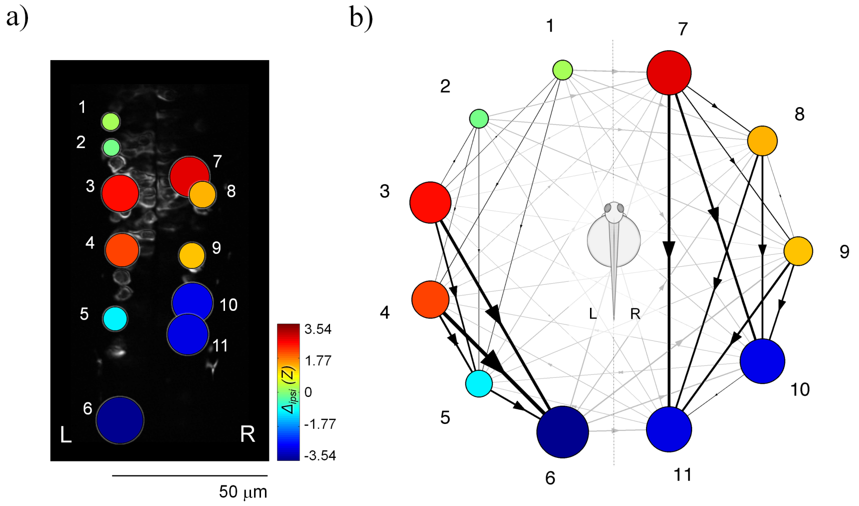


Figure 5. Rostro-caudal distribution of the nodal delta centrality in the representative zebrafish embryo. Panel a) The normalized ipsi value is represented for each node (motoneuron) as a colored circle superimposed on the field of view. The larger the circle, the more central is the node in terms of its tendency to act as a transmitter (red color, positive value) or receiver (blue color, negative value) hub of information flow. Panel b) The same normalized ipsi centrality values are here represented within the neuronal GC network. Statistically significant GC influences are illustrated as directed arrows. The thicker the arrow the stronger the GC value is. Inter-hemicord directed links are illustrated in gray color for the sake of simplicity.

More details in [9].

6.13. 2D harmonic filtering of MR phase images in multicenter clinical setting: towards a magnetic signature of cerebral microbleeds

Participants: Takoua Kaaouana [correspondant], Ludovic de Rochefort, Thomas Samaille, Nathalie Thiery, Carole Dufouil, Christine Delmaire, Didier Dormont, Marie Chupin.

Cerebral microbleeds (CMBs) have emerged as a new imaging marker of small vessel disease. Composed of hemosiderin, CMBs are paramagnetic and can be detected with MRI sequences sensitive to magnetic susceptibility (typically, gradient recalled echo T2* weighted images). Nevertheless, their identification remains challenging on T2* magnitude images because of confounding structures and lesions. In this context, T2* phase image may play a key role in better characterizing CMBs because of its direct relationship with local magnetic field variations due to magnetic susceptibility difference. To address this issue, susceptibility-based imaging techniques were proposed, such as Susceptibility Weighted Imaging (SWI) and Quantitative Susceptibility Mapping (QSM). But these techniques have not yet been validated for 2D clinical data in multicenter settings. Here, we introduce 2DHF, a fast 2D phase processing technique embedding both unwrapping and harmonic filtering designed for data acquired in 2D, even with slice-to-slice inconsistencies. This method results in internal field maps which reveal local field details due to magnetic inhomogeneity within the region of interest only. This technique is based on the physical properties of the induced magnetic field and should yield consistent results. A synthetic phantom was created for numerical simulations. It simulates paramagnetic and diamagnetic lesions within a "brain-like’ tissue, within a background. The method was evaluated on both this synthetic phantom and multicenter 2D datasets acquired in a standardized clinical setting, and compared with two state-of-the-art methods. It proved to yield consistent results on synthetic images and to be applicable and robust on patient data. As a proof-of-concept, we finally illustrate that it is possible to find a magnetic signature of CMBs and CMCs on internal field maps generated with 2DHF on 2D clinical datasets that gives consistent results with CT-scans in a subsample of 10 subjects acquired with both modalities. See Fig. 6

More details in [16].

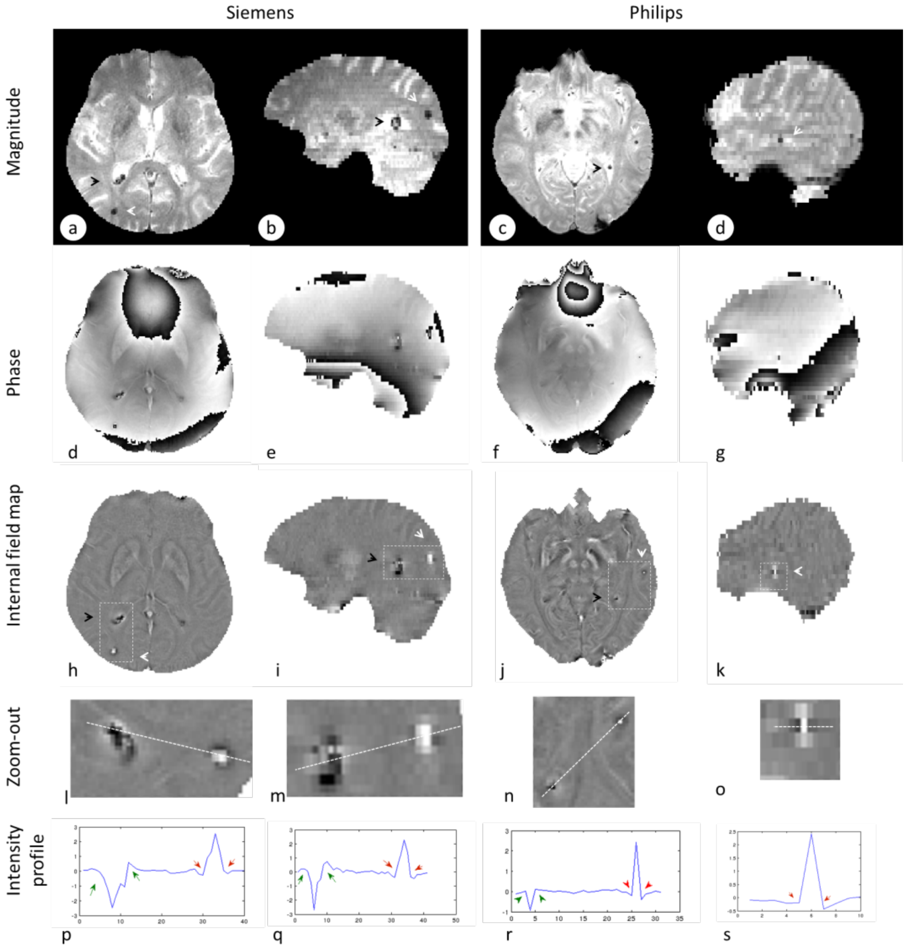


Figure 6. Siemens (left) and Philips (right) axial and sagittal views. Magnitude image (first row), native phase image (second row) and internal field map (third row). Fourth row shows a zoomed out region corresponding to the white rectangle showing CMB with a dipolar pattern (white arrow) and a physiologic calcification of the choroid plexus (black arrow). Note that panel l was rotated. A 1D intensity profile calculated through CMBs and calcification in the zoomed region is displayed in the last row. Note the intensity sign inversion for both side of CMBs (red arrow head), and the calcification (green arrow head). Double headed arrows on panels (l-o) indicate the location of the lines used to generate the intensity profiles.

ASCLEPIOS Project-Team

5. New Results

5.1. Highlights of the Year

- Nicholas Ayache was elected a member of the Académie des sciences on 18th Nov. 2014.
- Nicholas Ayache received the “Grand Prix Inria – Académie des sciences 2014“ for his major contributions to Informatics and Computational Sciences at Inria.
- Nicholas Ayache taught the "Personalized Digital Patient" course at the Collège de France on the annual chair "Informatics and Computational Sciences".
- Hervé Lombaert was awarded and ranked 1st in computer science at the highly selective NSERC Postdoctoral Fellowship (Top funding agency in Canada).
- Nina Miolane and Bishesh Khanal won the first prize in the “Popular Vote Awards” at the MICCAI 2014 Educational Challenge for their video on “Statistics on Lie groups for Computational Anatomy“.

BEST PAPER AWARD :

[12] **MICCAI Workshop on Abdominal Imaging – Computational and Clinical Applications**. C. AUDIGIER, T. MANSI, H. DELINGETTE, S. RAPAKA, V. MIHALEF, D. CARNEGIE, E. BOCTOR, M. CHOTI, A. KAMEN, D. COMANICIU, N. AYACHE.

5.2. Medical Image Analysis

5.2.1. 3D/2D Coronary Arteries Registration

Participants: Thomas Benseghir [correspondent], Grégoire Malandain [Morpheme Team], Régis Vaillant [GE-Healthcare], Nicholas Ayache.

This work has been performed in colaboration with GE-Healthcare (Buc) and the Morpheme team at Inria SAM.

3D/2D Registration, Computed Tomography Angiography, X-ray Fluoroscopy, Coronary Arteries, Vascular Tree

Integrating vessel calcifications and occlusion information, extracted from pre-operative 3D CT angiography images into a live fluoroscopic 2D image can greatly improve the guidance of percutaneous coronary interventions. Such task requires a registration step that must provide relevant correspondences between these two complementary modalities. We are developing a framework aiming at preserving the topology of the vascular structures matched between both images.

The introduction of topology in the pairing procedure allows to decrease the mismatches with respect to geometrically-based pairing procedures (e.g. Iterative Closest Point), which, in turn, improves the success rate of the registration method. This is exemplified by Fig. 3 where the proposed pairing method is compared to ICP.

5.2.2. Video Synchronization: An Approach Towards Endoscopic Re-localization

Participants: Anant Suraj Vemuri [correspondent], Nicholas Ayache.

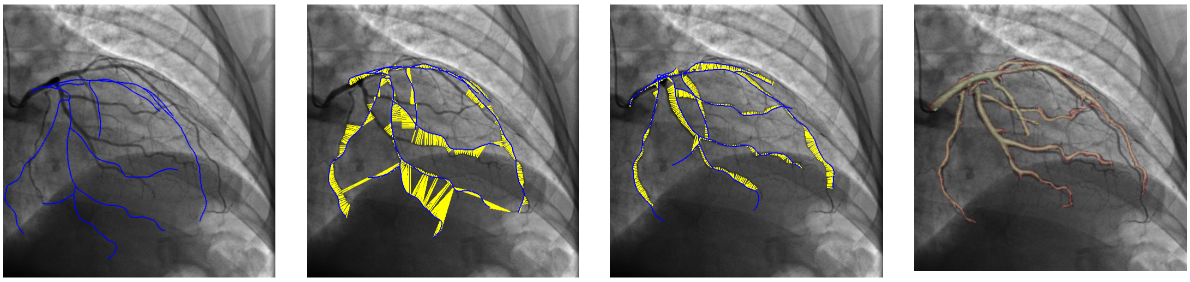


Figure 3. From left to right: 1. initial pose estimate (3D centerlines projection in blue); 2. iterative closest point algorithm registered position (point-pairings in yellow); 3: proposed method registered position; 4: resulting fusion between the two modalities

Endoscopy, Barrett's Esophagus, Re-localization, Electromagnetic Tracking

- Barrett's esophagus is the pre-malignant lesion for the majority of patients with esophageal adenocarcinoma. The evolution of the disease involves endoscopic surveillance for patients every 3-6 months, according to the Seattle protocol.
- The approach is labor-intensive and the primary problem is the inter-operative re-localization of these biopsy sites to guide the treatment.
- In an earlier work we had proposed a general framework for inter-operative biopsy site re-localization framework by introducing an Electro-magnetic tracking system (EMTS) into the loop and providing a way to inter-operatively register video sequences to provide a guided navigation in the esophagus.
- This work has been extended further to fit the operating room workflow. Two external landmarks have been added to the system setup as shown in Fig.4 , to obtain a complete reference frame with respect to the patient and to make the registration, patient specific [38]. The patient-localized reference frame allows the recovery of complete $SE(3)$ including the roll angle about the esophageal axis.

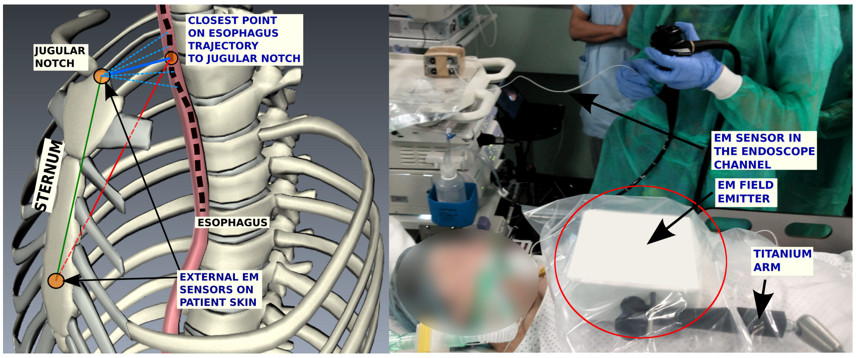


Figure 4. (Left) The orange circular markings indicate the position of the sensors taken as anatomical landmarks to form the reference frame for the patient. (Right) System setup in the operating room.

5.2.3. A sparse Bayesian framework for non-rigid registration

Participants: Loic Le Folgoc [correspondent], Hervé Delingette, Antonio Criminisi, Nicholas Ayache.

This work has been partly supported by Microsoft Research - Inria joint laboratory through its PhD Scholarship Programme and the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Registration, Automatic Relevance Determination, Uncertainty Quantification

We propose a sparse Bayesian framework for non-rigid registration. It provides a principled approach to efficiently find an optimal, sparse parameterization of deformations among any preset, widely overcomplete range of basis functions. It addresses open challenges in state-of-the-art registration, such as the automatic joint estimate of model parameters (e.g. noise and regularization levels). We have evaluated the feasibility and performance of our approach on cine MR, tagged MR and 3D US cardiac images, and show state-of-the-art results on benchmark datasets evaluating accuracy of motion and strain (see Fig.5). This work was presented during the MICCAI 2014 conference[20].

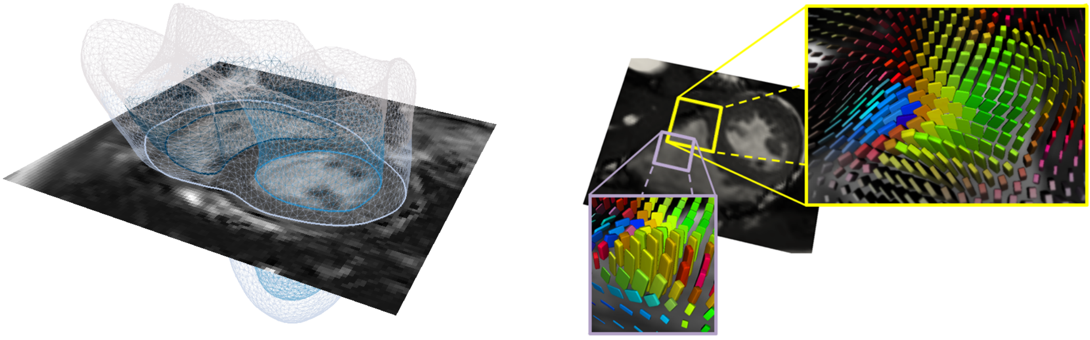


Figure 5. (Left) Mesh contour propagated to end systole via the registration output; (Right) Spatial uncertainty visualized as a tensor map.

5.2.4. Segmentation and anatomic variability of the cochlea and other temporal bone structures from medical images

Participants: Thomas Demarcy [correspondent], Hervé Delingette, Clair Vandersteen, Dan Gnansia [Oticon Medical], Nicholas Ayache.

This work is funded by a CIFRE grant involving Oticon Medical (Vallauris) and performed in collaboration with the IUFC Nice (Pr. Guevara) and CHU Nice (Pr. Raffaelli) .

image segmentation ; surgery planning ; shape modelling ; anatomic variability ; cochlear implant ; temporal bone

- We applied semi-automatic segmentation methods to extract anatomical structures on the inner ear on both micro-CT and CT scan images.
- μ-CT and CT images acquired on the same subject were fused with their segmentation.
- We designed a teaching tool[37] for advanced visualization of temporal bone structures (see Fig. 6).

5.2.5. Understanding cardiac planes of acquisition

Participants: Jan Margeta [correspondent], Nicholas Ayache, Daniel C Lee [Northwestern University], Antonio Criminisi [Microsoft Research Cambridge].

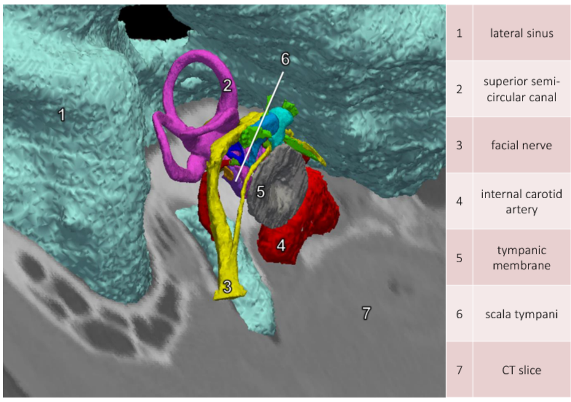


Figure 6. Virtual view of the posterior tympanotomy approach fused with a CT-scan

This work has been partly supported by Microsoft Research through its PhD Scholarship Programme, by ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images), and by the VP2HF FP7 research project.

Cardiac imaging, Machine learning, Magnetic resonance, Data wrangling

DICOM image format defines several tags by which the images can be queried and filtered. Many useful tags are however not standardized and must be cleaned prior to any large scale analysis.

- We developed a machine learning method for automatic recognition of cardiac planes of acquisition (See Fig. 7 for sample predictions).
- Our image based method achieved state of the art performance.
- This work was presented[23] at the Medical Image Understanding and Analysis conference in London.

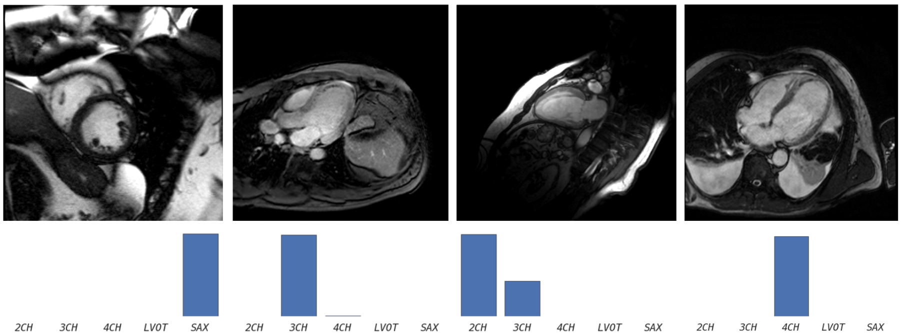


Figure 7. Examples of cardiac acquisition plane predictions and confidences

5.3. Computational Anatomy

5.3.1. Statistical Analysis of Diffusion Tensor Images of the Brain

Participants: Marco Lorenzi [Correspondent], Nicholas Ayache, Xavier Pennec.

Image non-linear registration, Longitudinal modeling, Alzheimer's disease

Alzheimer's disease is characterized by the co-occurrence of different phenomena, starting from the deposition of amyloid plaques and neurofibrillary tangles, to the progressive synaptic, neuronal and axonal damages. The brain atrophy is a sensitive marker of disease progression from pre-clinical to the pathological stages, and computational methods for the analysis of magnetic resonance images of the brain are currently used for group-wise (cross-sectional) and longitudinal studies of pathological morphological changes in clinical populations. The aim of this project is to develop robust and effective computational instruments for the analysis of longitudinal brain changes. In particular novel methods based on non-linear diffeomorphic registration have been investigated in order to reliably detect and statistically analyze pathological morphological changes [5] (see Fig.8). This project is also focused in the comparison of the trajectories of longitudinal morphological changes [31] estimated in different patients. This is a central topic for the development of statistical atlases of the longitudinal evolution of brain atrophy.

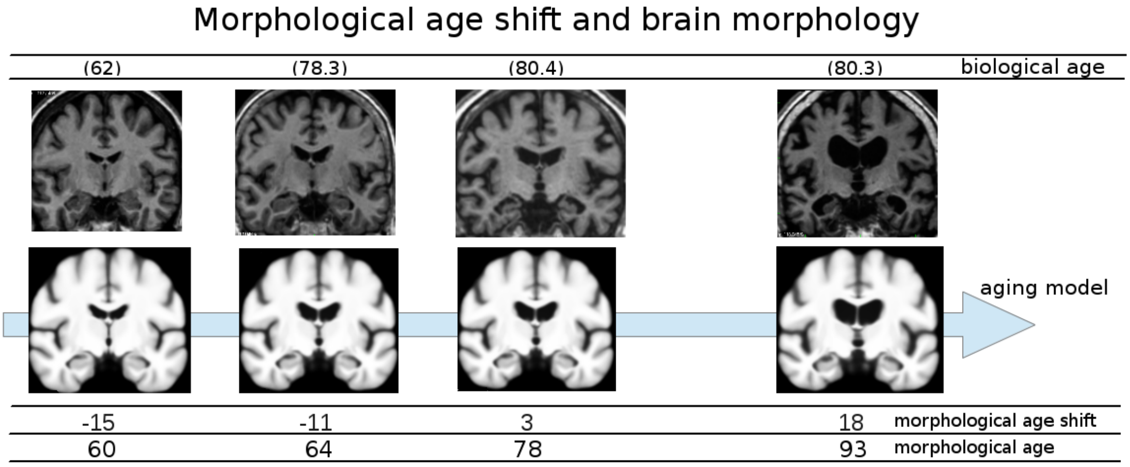


Figure 8. Modeled longitudinal brain changes in normal aging extrapolated from -15 to 18 years, and corresponding observed patient anatomies with estimated morphological age and age shift (biological age in parenthesis). Our modeling framework describes meaningful anatomical changes observed in clinical groups.

5.3.2. Statistical Learning via Synthesis of Medical Images

Participants: Hervé Lombaert [Correspondent], Nicholas Ayache, Antonio Criminisi.

This work has been partly supported by a grant from Microsoft Research-Inria Joint Centre, by ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images)

Statistical learning, Synthesis

Machine learning approaches typically require large training datasets in order to capture as much variability as possible. Application of conventional learning methods on medical images is difficult due to the large variability that exists among patients, pathologies, and image acquisitions. The project aims at exploring how realistic image synthesis could be used, and improve existing machine learning methods.

First year tackled the problem of better exploiting existing training sets, via a smart modeling of the image space (Fig. 9), and applying conventional random forests using guided bagging [21]. Synthesis of complex data, such as cardiac diffusion images (DTI), was also done. Synthesis of complex shapes, using spectral graph decompositions, is currently on-going work.

The modeling of shapes also includes novel representations based on the spectral decomposition of images[4] which are more robust to large deformations when comparing multiple patients.

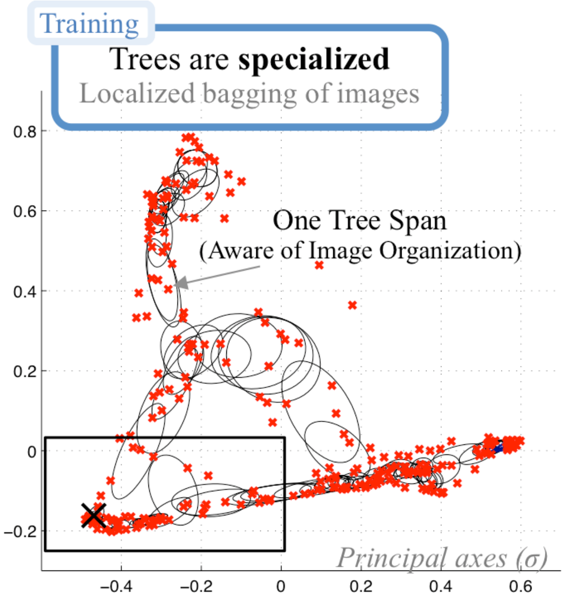


Figure 9. Laplacian Forest, where images are here represented as points, and where decision trees are trained using the spatial organization of these images on a reduced space.

5.3.3. Statistical analysis of heart shape, deformation and motion

Participants: Marc-Michel Rohé [correspondent], Xavier Pennec, Maxime Sermesant.

This work was partly supported by the FP7 European project MD-Paedigree and by ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images)

Statistical analysis, Registration, Reduced order models, Machine learning

The work aims at developping statistical tools to analyse cardiac shape, deformation, and motion. In particular, we are interested in developping reduced order models so that the variability within a population described by a complex model can be reduced into few parameters or modes that are clinically relevant. We use these modes to represent the variability seen in a population and to relate this variability with clinical parameters, and we build group-wise statistics which relate these modes to a given pathology. We focus on cardiomyopathies and the cardiovascular disease risk in obese children and adolescents.

5.3.4. Geometric statistics for Computational Anatomy

Participants: Nina Miolane [Correspondent], Xavier Pennec.

Lie groups, pseudo-Riemannian, Statistics, Computational Anatomy

Lie groups are widely used in mathematical models for Medical Imaging. In Computational Anatomy for example, an organ's shape can be modeled as the deformation of a reference shape, in other words : as an element of a Lie group. If one wants to analyze the variability of the human anatomy, e.g. to help diagnose diseases, one has to perform statistics on Lie groups. We investigate the geometric structures on Lie groups that enable to define consistent statistics. A Lie group G is a manifold with an additional group structure. Statistics on *Riemannian* manifolds have be studied throughout the past years. One may wonder if we could use the theory of statistics on Riemannian manifolds for statistics on G. To this aim, we need to define a Riemannian metric on the Lie group that is *compatible with the group structure*: a so-called *bi-invariant* metric. However, it is known that most Lie groups do not admit any bi-invariant metric. One may wonder if we could generalize the theory of statistics on Riemannian manifolds to pseudo-Riemannian manifolds and use it for statistics on G. To this aim, we need to define a bi-invariant pseudo-metric on G. How many Lie groups do admit such a pseudo-metric and can we compute it? These investigations and their results (see Fig. 10) were presented at MaxEnt 2014 [24].

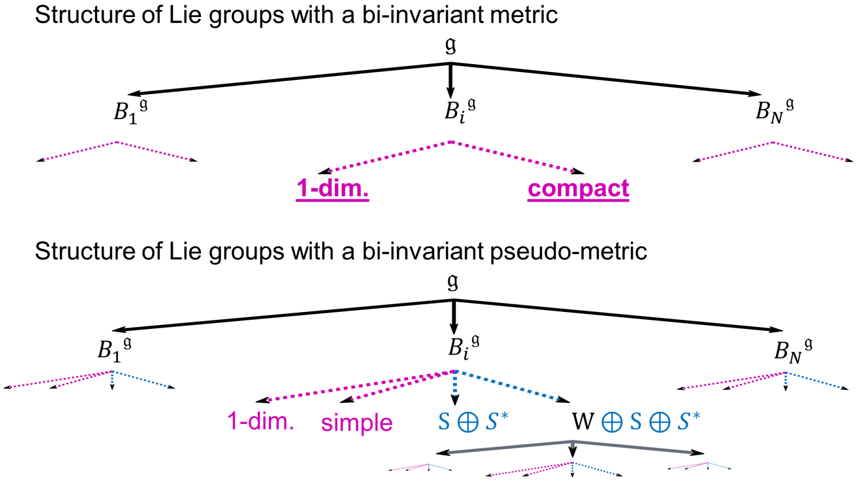


Figure 10. Structure of Lie groups on which one can define a bi-invariant metric or a bi-invariant pseudo-metric. The black levels of the tree represent the adjoint decomposition of the Lie algebra, the dashed lines represent the possible algebraic types of the substructures. Note the recursive construction in the pseudo-Riemannian case.

5.3.5. Statistical Analysis of Diffusion Tensor Images of the Brain

Participants: Vikash Gupta [correspondent], Nicholas Ayache, Xavier Pennec.

Population specific multimodal brain atlas for statistical analysis of white matter tracts on clinical DTI.

HIV virus can cross the hematoencephallic barrier and affect the neural connectivity in the human brain causing compromised motor controls, loss in episodic, long term memory and working memory, loss in attention/concentration and visual agnosia. These cognitive losses are characterized by the neuropsychological (NP) test scores and believed to be correlated with destruction of white matter (WM) integrity among the HIV patients. For quantifying the loss in WM integrity, the HIV subjects are compared against controls using a tract based spatial statistics (TBSS) routine. The standard TBSS routines uses univariate statistics using the fractional anisotropy (FA) maps. However, we improved on the existing routines using tensor based registration for normalizing the diffusion tensor images (DTI) followed by a multivariate statistics using the full tensor information. With the improved method it is possible to detect differences in WM regions which was not possible using the existing TBSS routines. For this study a population specific multimodal (T1 and DTI) brain atlas was developed from the population. The joint atlas also contains a probabilistic parcellation of WM regions in the brain which can be used for region of interest (ROI) based statistical studies (see Fig.11).

5.3.6. Longitudinal Analysis and Modeling of Brain Development

Participants: Mehdi Hadj Hamou [correspondent], Xavier Pennec, Nicholas Ayache.

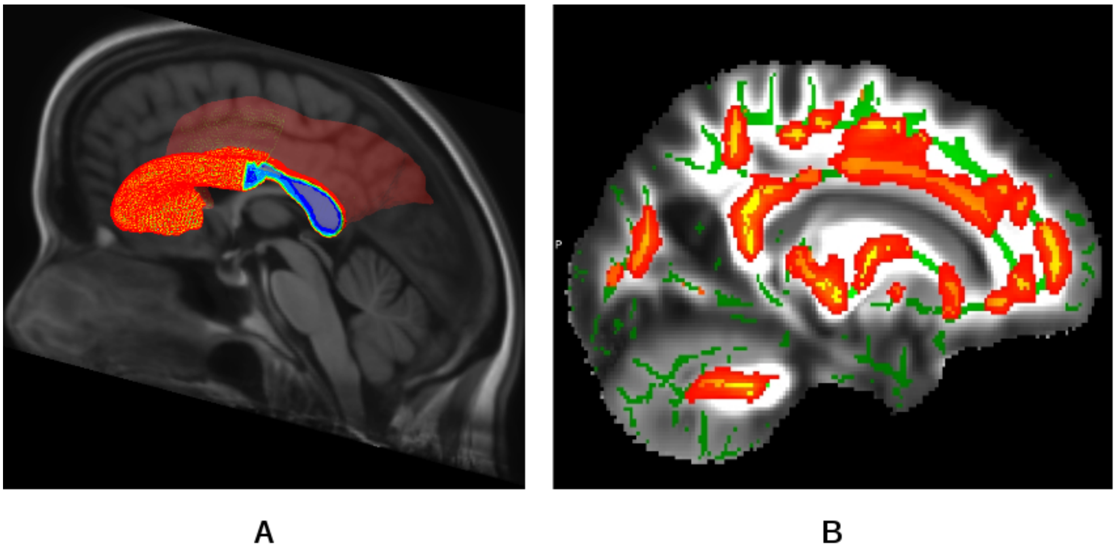


*Figure 11. **A:** Probabilistic parcellation of corpus callosum with blue and red being the maximum and minimum probability regions respectively. **B:** Multivariate statistics on white matter tracts. The red-yellow sections show statistically significant differences*

This work is partly funded through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Brain development, adolescence, longitudinal analysis, non-rigid registration algorithm, extrapolation, interpolation

This work is divided into 2 complementary studies about longitudinal trajectories modeling:

- Diffeomorphic registration parametrized by Stationary Velocity Fields (SVF) is a promising tool already applied to model longitudinal changes in Alzheimer's disease. However, the validity of these model assumptions in faithfully describing the observed anatomical evolution needs to be further investigated. In this work, we thus analyzed the effectiveness of linear regression of SVFs in describing anatomical deformations estimated from past and future observations of the MRIs.
- Due to the lack of tools to capture the subtle changes in the brain, little is known about its development during adolescence. The aim of this project is to provide quantification and models of brain development during adolescence based on diffeomorphic registration parametrized by SVFs (see Fig.12). We particularly focused our study on the link between gender and the longitudinal evolution of the brain. This work was done in collaboration with J.L. Martinot et H. Lemaître (Inserm U1000).

5.4. Computational Physiology

5.4.1. Biophysical Modeling and Simulation of Longitudinal Brain MRIs with Atrophy in Alzheimer's Disease

Participants: Bishesh Khanal [correspondent], Nicholas Ayache, Xavier Pennec.

This work has been partly supported by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

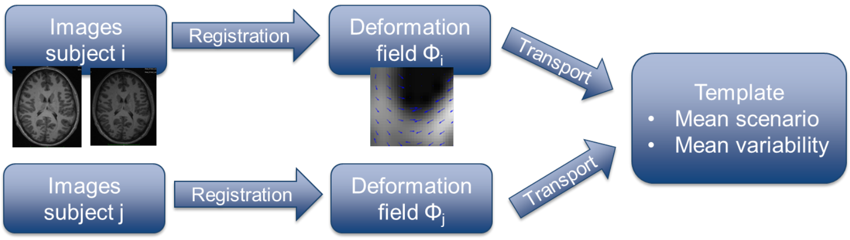


Figure 12. Pipeline for the longitudinal analysis of brain development during adolescence.

Alzheimer's Disease (AD), modeling brain deformation, biophysical model, simulation

- We propose a biophysical model of brain deformation due to atrophy in Alzheimer's Disease(AD) [17]. The model allows simulation of longitudinal brain MRIs with a desired level of atrophy in brain parenchyma. Here we enhanced our previous implementation to model brain parenchyma and cerebrospinal fluid (CSF) differently so that there is no need to prescribe atrophy in CSF region (see Fig.13).
- The model could be used to explore different possible hypotheses about evolution of atrophy in the brain and how it affects the brain shape changes.

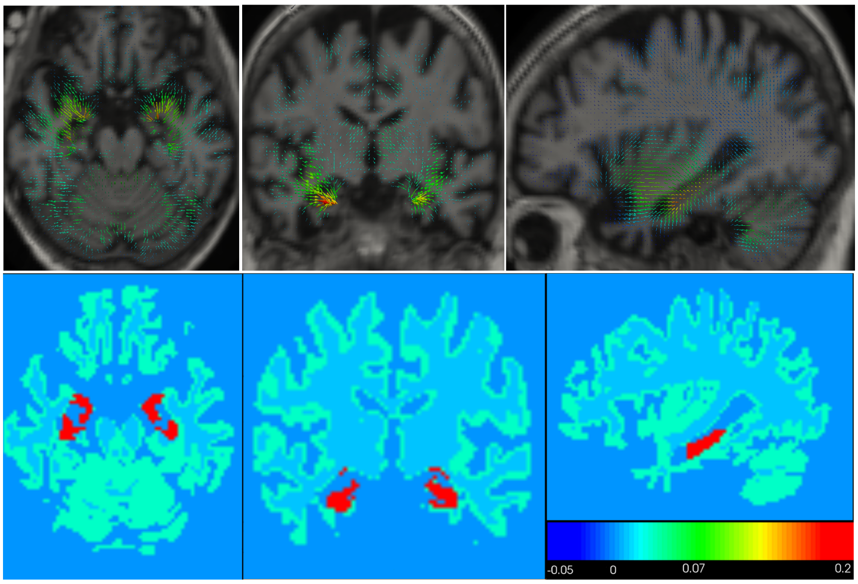


Figure 13. An example of obtained deformation field (top) from the model for the prescribed atrophy (bottom). From left to right: Axial, Coronal and Sagittal views.

5.4.2. Glioblastoma : Study of the vasogenic edema

Participants: Matthieu Lê [correspondent], Hervé Delingette, Jan Unkelbach [Massachussetts General Hospital], Nicholas Ayache.

This work is carried out between Ascelpios research group, Inria Sophia Antipolis, France and the Department of Radiation Oncology of the Massachusetts General Hospital, Boston, USA.

Glioblastoma, Vasogenic Edema, Radiotherapy, Target Delineation

- We studied the impact of anti-angiogenic treatment on the MRI appearance of glioblastoma.
- We studied how MRI extracted features could help distinguish between the vasogenic edema and the tumor infiltration[22].
- We analyzed the impact of excluding the vasogenic edema from the gross tumor volume during radiation therapy (see Fig. 14). Our approach leads to a dose more comformal to the underlying tumor cell density knowing that prescribing less dose might open the way for later re-irradiation.

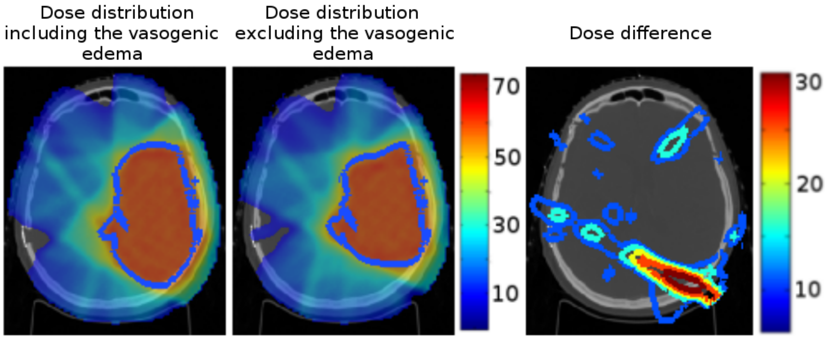


Figure 14. Comparison of the dose distribution including the vasogenic edema (clinical practice) and the excluding the estimated vasogenic edema (proposed method).

5.4.3. Image-based Prediction of Cardiac Ablation Targets

Participants: Rocio Cabrera Lozoya [correspondent], Maxime Sermesant, Nicholas Ayache.

Electrophysiology, ablation planning, machine learning

Ventricular radio-frequency ablation (RFA) can have a critical impact on preventing sudden cardiac arrest but is challenging due to a highly complex arrhythmogenic substrate. We aim at identifying local image characteristics capable of predicting the presence of local abnormal ventricular activities (LAVA). This could lead to pre-operatively and non-invasively improve and accelerate the procedure.

- We present the use of intensity and texture-based local imaging features in the vicinity of myocardial scar and grey zones towards the prediction of RFA target localisation (see Fig.15).
- We detail the uncertainty in the data and explore its impact on the classification results.
- A preliminary output with visual interpretation and potential use in a clinical environment was presented.
- The encouraging obtained results warrant further investigation and open up possibilities for non-invasive cardiac arrhythmia ablation planning. [13]

5.4.4. Personalised Canine Electromechanical Model of the Heart

Participants: Sophie Giffard-Roisin [correspondent], Maxime Sermesant, Hervé Delingette, Stéphanie Marchesseau, Nicholas Ayache.

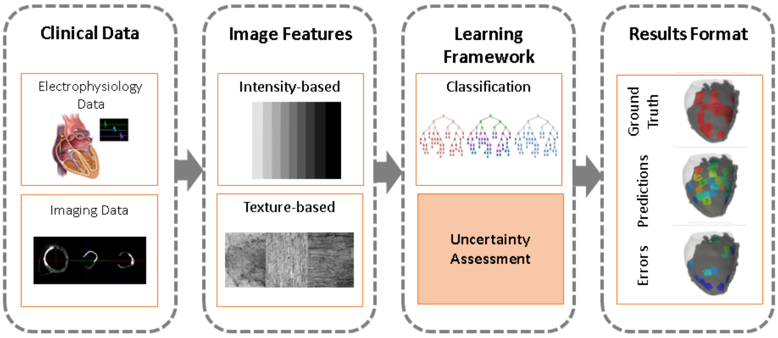


Figure 15. Pipeline showing the processing of our multimodal data. It includes an image feature extraction phase, followed by a classification with uncertainty assessment stage. The rightmost panel shows the preliminary output result format for a clinical environment.

This work has been supported by the European Project FP7 under grant agreement VP2HF (no 611823) and the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Cardiac Modelling, Personalised Simulation, Electrical and Mechanical Simulation

- We studied the coupled electro-mechanical modelling of the heart, where the mechanics is handled by the Bestel-Clement-Sorine model while the electrophysiological phenomena is driven by an Eikonal model (see Fig. 16).
- We participated to the STACOM'2014 LV Mechanics Challenge in Boston[15] where four healthy canine clinical data (left ventricles) were provided. The validation was performed on local displacements. Our model has been calibrated by a quantitative sensitivity study as well as a personalized automatic calibration.

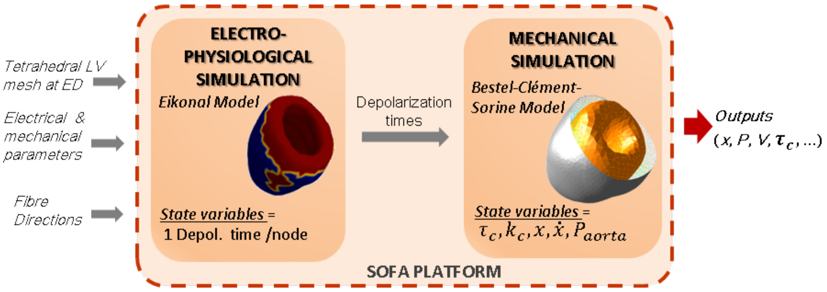


Figure 16. Complete electromechanical pipeline used for the simulations of four healthy canine hearts.

5.4.5. Computational modeling of radiofrequency ablation for the planning and guidance of abdominal tumor treatment

Participants: Chloé Audigier [correspondent], Hervé Delingette, Tommaso Mansi, Nicholas Ayache.

This PhD is carried out between Asclepios research group, Inria Sophia Antipolis and the Image Analytics and Informatics global field, Siemens Corporate Research, Princeton, USA.

Radio Frequency Abation, Patient-Specific Simulation, Lattice Boltzmann Method, Computational Fluid Dynamics, Heat Transfer, Therapy Planning, Liver

Radiofrequency abation (RFA) is a minimally invasive therapy suited for liver tumor ablation. However a patient-specific predicitive tool to plan and guide the treatment is required. We developed a computational framework for patient-specific planning of RFA (see Fig.17) :

- a personalised forward model of RFA:
- A patient-specific detailed anatomical model of the liver is estimated from standard CT image and meshed to generate a tetrahedral volume mesh. The structures of interest include the parenchyma, lesion, hepatic vein and vena cava.
- A Computation Fluic Dynamic and porous media solver using the Lattice Boltzmann Method is used to compute the patient-specific blood flow in the hepatic circulatory system and the blood flow distribution inside the parenchyma.
- Bio-heat equation has been implemented with a Lattice Boltmann Method also to model efficiently the heat propagation in biological tissues accounting for the cooling effect of neighboring vessels. A cell death model have been combined to account for the cellular necrosis.

Then this forward model is used to estimate patient-specific model parameters as presented in the ABDI workshop at MICCAI 2014 [12]. This work presented obtained the best paper award of the workshop.

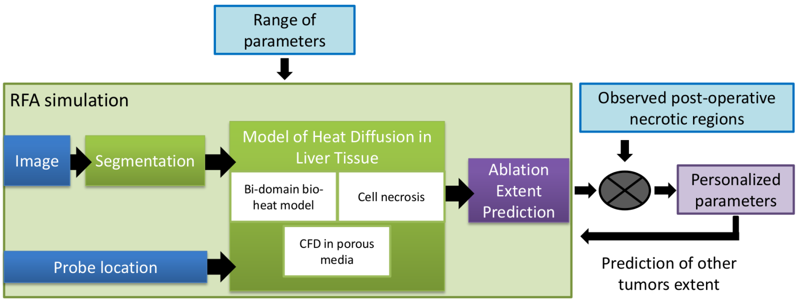


Figure 17. Steps of the proposed method for parameters personnalisation (blue: input, green: processes, purple: output).

5.4.6. Multi-channel patch-based glioma segmentation

Participants: Nicolas Cordier [correspondent], Hervé Delingette, Nicholas Ayache.

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Brain, MRI, Glioma, Patch-based Segmentation, Tumor Simulation

The segmentation of glioblastoma, the most severe case of brain tumors, is a crucial step for diagnostic assessment and therapy planning. In order to perform the manual delineation of the tumor compartments, the clinicians have to concurrently screen multi-channel 3D MRI, which makes the process both time-consuming and subject to inter-expert delineation variability. We are building upon the patch-based segmentation framework, the state-of-the-art for the segmentation of healthy brain structures, to present automatic glioma segmentation algorithms. Our 2013 submission to the MICCAI Brain Tumor Segmentation Challenge has been improved by:

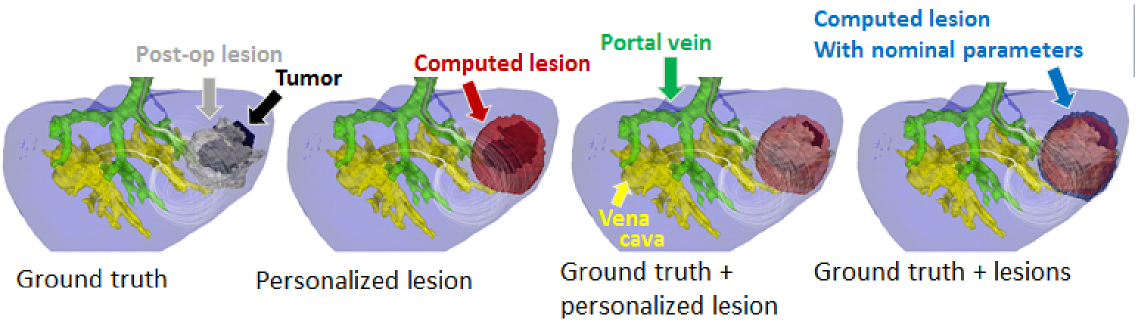


Figure 18. Predicted necrosis compared qualitatively well with ground truth (necrosis zone observed on a post-operative image).

- replacing the heuristic label fusion strategy with a more robust approach,
- integrating information such as statistics of appearance and position,
- generating configurations of synthetic training patches,
- filtering out the training patches for which the labels are less reliable (see Fig.19).

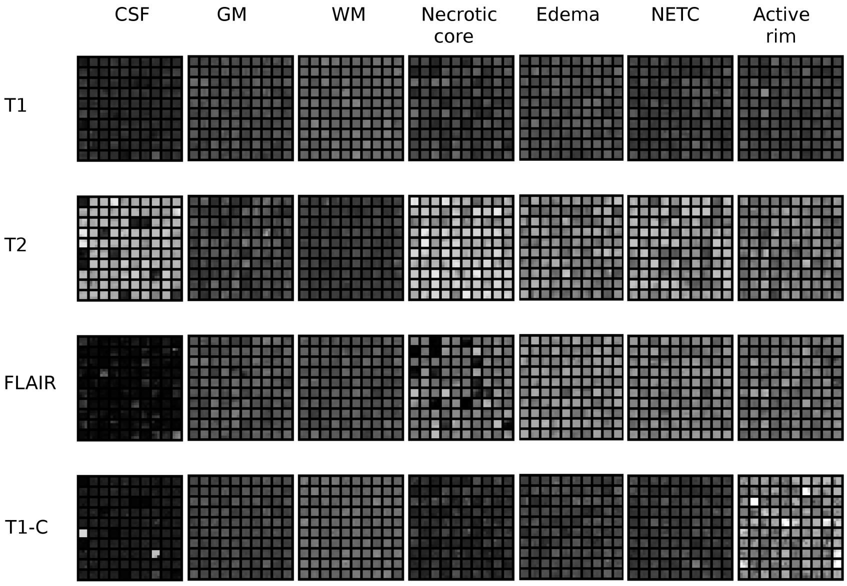


Figure 19. 2D slices of randomly sampled 3D multi-channel patches. From left to right: the different MR channels (T1, T2, T2-FLAIR, contrast enhanced T1). From top to bottom: cerbrospinal fluid (CSF), grey matter (GM), white matter (WM), necrotic tumor core, edema, non-enhancing tumor core (NETC), and active rim.

ATHENA Project-Team

6. New Results

6.1. Highlights of the Year

Maureen Clerc was awarded the PIERRE FAURRE Prize by the French Academy of Sciences. This award recognizes her outstanding contributions to the modelling and interpretation of electrical signals in the brain. The ceremony took place at the Institut de France on October 14th, 2013.

Emmanuel Caruyer was awarded the AFRIF Best PhD thesis award 2013 for his work "Q-space diffusion MRI: Acquisition and Signal Processing" performed under the direction of Rachid Deriche. He received the award thesis AFRIF 2013 during RFIA Conference held from June 30 to July 4, 2014 in Rouen.

Rachid Deriche was awarded the title of Honorary Doctor (honoris causa) from the University of Sherbrooke, Canada. This award recognises his achievements and contributions to image processing, computer vision and computational brain imaging. The title was awarded at the academic conferment ceremony held on September 20th, 2014 at the University of Sherbrooke.

Théo Papadopoulo has been promoted to the position of Research Director Class 2, starting from October 1st, 2014.

6.2. Modeling in Diffusion MRI

6.2.1. Non-Negative Spherical Deconvolution (NNSD) for estimation of fiber Orientation Distribution Function in single-/multi-shell diffusion MRI

Participants: Jian Cheng [University of North Carolina at Chapel Hill,USA], Tianzi Jiang [LIAMA, China], Shen Dinggang [University of North Carolina at Chapel Hill,USA], Yap Pew-Thian [University of North Carolina at Chapel Hill,USA], Rachid Deriche.

Spherical Deconvolution (SD) is commonly used for estimating fiber Orientation Distribution Functions (fODFs) from diffusion-weighted signals. Existing SD methods can be classified into two categories: 1) Continuous Representation based SD (CR-SD), where typically Spherical Harmonic (SH) representation is used for convenient analytical solutions, and 2) Discrete Representation based SD (DR-SD), where the signal profile is represented by a discrete set of basis functions uniformly oriented on the unit sphere. A feasible fODF should be non-negative and should integrate to unity throughout the unit sphere SS2. However, to our knowledge, most existing SH-based SD methods enforce non-negativity only on discretized points and not the whole continuum of SS2. Maximum Entropy SD (MESD) and Cartesian Tensor Fiber Orientation Distributions (CT-FOD) are the only SD methods that ensure non-negativity throughout the unit sphere. They are however computational intensive and are susceptible to errors caused by numerical spherical integration. Existing SD methods are also known to overestimate the number of fiber directions, especially in regions with low anisotropy. DR-SD introduces additional error in peak detection owing to the angular discretization of the unit sphere. This work proposes a SD framework, called Non-Negative SD (NNSD), to overcome all the limitations above. NNSD is significantly less susceptible to the false-positive peaks, uses SH representation for efficient analytical spherical deconvolution, and allows accurate peak detection throughout the whole unit sphere. We further show that NNSD and most existing SD methods can be extended to work on multi-shell data by introducing a three-dimensional fiber response function. We evaluated NNSD in comparison with Constrained SD (CSD), a quadratic programming variant of CSD, MESD, and an L1-norm regularized non-negative least-squares DR-SD. Experiments on synthetic and real single-/multi-shell data indicate that NNSD improves estimation performance in terms of mean difference of angles, peak detection consistency, and anisotropy contrast between isotropic and anisotropic regions.

This work has been published in [11].

6.2.2. *Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI*

Participants: Alessandro Daducci [LTS5, Ecole Polytech. Fed. de Lausanne (EPFL)], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Emmanuel Caruyer, Rachid Deriche.

Validation is arguably the bottleneck in the diffusion magnetic resonance imaging (MRI) community. This work evaluates and compares 20 algorithms for recovering the local intra-voxel fiber structure from diffusion MRI data and is based on the results of the "HARDI reconstruction challenge" organized in the context of the "ISBI 2012" conference. Evaluated methods encompass a mixture of classical techniques well known in the literature such as diffusion tensor, Q-Ball and diffusion spectrum imaging, algorithms inspired by the recent theory of compressed sensing and also brand new approaches proposed for the first time at this contest. To quantitatively compare the methods under controlled conditions, two datasets with known ground-truth were synthetically generated and two main criteria were used to evaluate the quality of the reconstructions in every voxel: correct assessment of the number of fiber populations and angular accuracy in their orientation. This comparative study investigates the behavior of every algorithm with varying experimental conditions and highlights strengths and weaknesses of each approach. This information can be useful not only for enhancing current algorithms and develop the next generation of reconstruction methods, but also to assist physicians in the choice of the most adequate technique for their studies.

This work has been published in [12]

6.2.3. *Comparison of sampling strategies and sparsifying transforms to improve compressed sensing diffusion spectrum imaging*

Participants: Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Guillaume Gilbert [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion Spectrum Imaging enables to reconstruct the ensemble average propagator (EAP) at the expense of having to acquire a large number of measurements. Compressive sensing offers an efficient way to decrease the required number of measurements. The purpose of this work is to perform a thorough experimental comparison of three sampling strategies and six sparsifying transforms to show their impact when applied to accelerate compressive sensing-diffusion spectrum imaging. In this work, we propose a novel sampling scheme that assures uniform angular and random radial q-space samples. We also compare and implement six discrete sparse representations of the EAP and thoroughly evaluate them on synthetic and real data using metrics from the full EAP, kurtosis, and orientation distribution function. The discrete wavelet transform with Cohen–Daubechies–Feauveau 9/7 wavelets and uniform angular sampling in combination with random radial sampling showed to be better than other tested techniques to accurately reconstruct the EAP and its features. In conclusion, it is important to jointly optimize the sampling scheme and the sparsifying transform to obtain accelerated compressive sensing-diffusion spectrum imaging. Experiments on synthetic and real human brain data show that one can robustly recover both radial and angular EAP features while undersampling the acquisition to 64 measurements (undersampling factor of 4).

This work has been published in [16]

6.2.4. *DSI 101: Better ODFs for Free*

Participants: Michael Paquette [SCIL Lab., Sherbrooke University], Sylvain Merlet, Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion Spectrum Imaging (DSI) is a well established method to recover the Ensemble Average Propagator (EAP). The orientation distribution function (ODF) is computed from this discretized EAP and used for tractography. However, there are several important implementation considerations that are tossed aside in the literature and the publicly available softwares. We investigate all the real steps necessary to go from the DSI signal to the ODF and provide applicable recommendations that greatly improve the accuracy of the local orientation detected. These recommendations come ”free-of-charge” as they are applicable to all existing DSI data and do not require a significant increase in computation time.

This work has been published in [26]

6.2.5. Comparison between discrete and continuous propagator indices from Cartesian q-space DSI sampling

Participants: Mauro Zucchelli [Dpt of Computer Science, University of Verona], Eleftherios Garyfallidis [SCIL Lab., Sherbrooke University], Michael Paquette [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Sylvain Merlet.

DSI is often considered the state-of-the-art technique to analyze q-space measurements sampled from a Cartesian grid. The 3D fast Fourier transform is used to directly obtain a discrete version of the EAP (Ensemble Average Propagator). DSI was one of the first techniques used to infer complex fiber configurations as it allows resolving crossings. In principle, DSI also captures some radial information which,in theory, can be used to extract diffusion features of the EAP. However, a discrete propagator representation suffers from a limited frequency band, which makes infinite integration impossible. Hence, EAP derived indices 2,3 are problematic and quantitatively questionable, as one needs to artificially normalize and approximate the infinite integrals. Combined with the recent popularity of DSI in the Human Connectome Project, it is important to investigate the different angular and EAP indices that can be computed from these DSI datasets. In this work, we investigate alternatives to the discrete model-free approach of DSI and investigate the Simple Harmonic Oscillator based Reconstruction and Estimation 3 (SHORE) models based on the evaluation of (i) the orientation distri bution function (ODF) ; (ii) the return to the origin probability (RTOP) and (iii) the mean square displacement (MSD).

This work has been published in [33]

6.2.6. Odf Maxima Computation Using Hill Climbing Algorithm

Participants: Makhlouf Laouchedi [USTHB, Algeria], Thinhinane Megherbi [USTHB, Algeria], Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

Methods like Diffusion Spectrum Imaging (DSI), High Angular Resolution Diffusion Imaging (HARDI) and the High Order Tensor techniques have been proposed to reconstruct specific functions like the Orientation Distribution Function (ODF) whose maxima correspond to the directions of the multiple fibers. In this work, we are interested to extract all the crossing fibers characterized as the maxima of the Orientation Distribution Function (ODF). A Hill Climbing algorithm based approach has been developed and implemented to efficiently and accurately extract all the fibers. Promising experimental results obtained with synthetic and real data illustrate the potential of the technique.

This work has been published in [24]

6.2.7. Greedy NNLS: Fiber Orientation Distribution from Non-Negatively Constrained Sparse Recovery

Participants: Aurobrata Ghosh, Rachid Deriche.

In this work, we validated experimentally the merits of the Non-Negative Least Squares (NNLS) for the constrained sparse recovery of the Fiber Orientation Distribution (FOD) and compared it with classical l1-minimization. The FOD is a robust model for mapping crossing white matter fibers. However, its angular resolution depends on the spherical harmonic basis order, which can imply a large number of acquisitions. Further, it is necessary to compute the maxima of the FOD to derive the fiber directions. It is possible to kill the two proverbial birds with a single stone by using a non-negatively constrained sparse recovery FOD estimation with NNLS.

From our experiments, we confirmed results from literature to show that NNLS converges to highly sparse solutions which are correctly constrained, while l1-minimization is less sparse, contains negative solutions and is unstable with noisy data. Finally, we discussed the NLS algorithm and attributed the sparsity to its design, which mirrors the design of Orthogonal Matching Pursuit (OMP)

This work has been published in [22]

6.2.8. Crossing Fibers Detection with an Analytical High Order Tensor Decomposition

Participants: Thinhinane Megherbi [USTHB, Algeria], Mouloud Kachouane [USTHB, Algeria], Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

Diffusion magnetic resonance imaging (dMRI) is the only technique to probe in vivo and noninvasively the fiber structure of human brain white matter. Detecting the crossing of neuronal fibers remains an exciting challenge with an important impact in tractography. In this, we tackle this challenging problem and propose an original and efficient technique to extract all crossing fibers from diffusion signals. To this end, we start by estimating, from the dMRI signal, the so-called Cartesian tensor fiber orientation distribution (CT-FOD) function, whose maxima correspond exactly to the orientations of the fibers. The fourth order symmetric positive definite tensor that represents the CT-FOD is then analytically decomposed via the application of a new theoretical approach and this decomposition is used to accurately extract all the fibers orientations. Our proposed high order tensor decomposition based approach is minimal and allows recovering the whole crossing fibers without any a priori information on the total number of fibers. Various experiments performed on noisy synthetic data, on phantom diffusion, data and on human brain data validate our approach and clearly demonstrate that it is efficient, robust to noise and performs favorably in terms of angular resolution and accuracy when compared to some classical and state-of-the-art approaches.

This work has been published in [15] and [34].

6.2.9. Complete set of Invariants of a 4th order tensor: the 12 tasks of HARDI from Ternary Quartics

Participants: Théodore Papadopoulo, Auro Ghosh, Rachid Deriche.

In this work, we presented a simple and systematic method to compute a functionally complete set of invariants of a non-negative 3D 4th order tensor with respect to 3D rotations. Intuitively, this transforms the tensor's non-unique ternary quartic (TQ) decomposition (from Hilbert's theorem) to a unique canonical representation independent of orientation.

Invariants play a crucial role in diffusion MRI. In DTI (2nd order tensors), invariant scalars (FA, MD...)have been successfully used in clinical applications. But DTI has limiations and HARDI models (e.g. 4th order tensors) have been proposed instead. These, however, lack invariant features and computing them systematically is challenging.

The invariants we propose, can be computed from two simple reduction steps, which first reduce an orthogonal class and then a rotation transform class of equivalent representations from the TQ coefficients. The resulting invariants are, by construction, (1) functionally complete, (2) functionally irreducible (if desired), (3) computationally efficient and (4) reversible – or mappable to the TQ coefficients or shape. These were the novelties of our contribution in comparison to prior work.

This work has been published in [25]

6.2.10. Fiber Orientation Distribution from Non-Negative Sparse Recovery

Participants: Thinhinane Megherbi [USTHB, Algeria], Auro Ghosh, Linda Oulebsir-Boumghar [USTHB, Algeria], Rachid Deriche.

In this work, we tested our non-negatively constrained sparse recovery algorithm for estimating the FOD on single shell phantom data provided by the ISBI'2014 challenge. We used the NNLS algorithm to estimate high order FODs (24th order) from just 20, 30 and 60 gradient directions and for various b-values of 1000, 2000, and 3000.

From the results, which are yet to be published, but can be viewed in their preliminary form online, it is clear that amongst the single shell algorithms, ours was good at fitting the signal and estimating the number of compartments. It performed well even with as low as 20 gradient acquisitions. Its major shortcoming was in underestimating the crossing angle and this needs to be improved upon.

This work has been published in [35]

6.2.11. How to get more out of a clinically feasable 64 gradient dMRI acquisition: Multi-Shell versus Single-Shell

Participants: Rutger H.j Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

For clinical applications the number of diffusion MRI (dMRI) samples that can be obtained is often limited by scanner time and patient comfort. For this reason one often uses short scanning protocols that acquire just 32 or 64 gradient directions using a single b-value to obtain diffusion measures such as the fractional anisotropy from Diffusion Tensor Imaging (DTI) or to estimate the white matter orientation using Constrained Spherical Deconvolution (CSD). Using 3D-SHORE and MAP-MRI, we show that by spreading the same number of dMRI samples over different b-shells (sampling angularly and radially) we can estimate not only the directionality of the white matter using the ODF, but also the radially dependent higher order diffusion measures that SHORE and MAP-MRI provide. This approach lends itself well for situations where acquisition time is limited, and is therefore particularly well suited for clinical applications.

This work has been submitted to ISMRM'2015.

6.3. From dMRI to Fiber Pathways

6.3.1. Towards quantitative connectivity analysis: reducing tractography biases

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion MRI tractography is often used to estimate structural connections between brain areas and there is a fast-growing interest in quantifying these connections based on their position, shape, size and length. However, a portion of the connections reconstructed with tractography is biased by their position, shape, size and length. Thus, connections reconstructed are not equally distributed in all white matter bundles. Quantitative measures of connectivity based on the streamline distribution in the brain such as streamline count (density), average length and spatial extent (volume) are biased by erroneous streamlines produced by tractography algorithms. In this work, solutions are proposed to reduce biases in the streamline distribution. First, we propose to optimize tractography parameters in terms of connectivity. Then, we propose to relax the tractography stopping criterion with a novel probabilistic stopping criterion and a particle filtering method, both based on tissue partial volume estimation maps calculated from a T1-weighted image. We show that optimizing tractography parameters, stopping and seeding strategies can reduce the biases in position, shape, size and length of the streamline distribution. These tractography biases are quantitatively reported using in-vivo and synthetic data. This is a critical step towards producing tractography results for quantitative structural connectivity analysis.

This work has been published in [13]

6.3.2. Choosing tractography parameters to improve connectivity mapping

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion-weighted imaging (DWI) is often used as a starting point for in vivo white matter (WM) connectivity to reconstruct potential WM pathways between brain areas. Tractography algorithms have many parameters which can influence reconstruction and connectivity. Various choices of parameters have been proposed. But how to choose the best set of parameters ? In this study, we varied three critical parameters while monitoring connectivity score using the Tractometer evaluation system on the International Symposium on Biomedical Imaging (ISBI) Challenge synthetic dataset. The three parameters were: The maximum deviation angle between two consecutive tractography steps (this addresses the hypothesis of smoothness of the WM pathways), the spherical function (SF) threshold (this aims at removing noisy propagation directions during the tractography process) and the initial SF threshold (this aims at removing initial noise at the seeds and to start tractography in a good tangent direction to the WM bundle).

This work has been published in [20]

6.3.3. Connectivity directionally-encoded color map: a streamline-based color mapping

Participants: Gabriel Girard [SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

In this work, we provide a novel method to map streamlines in a color image, which can be generated from any set of streamlines. We show that this novel orientation color-coded map based on streamline tractography can improve connectivity analysis.

This work has been published in [23]

6.4. From dMRI to Microstructures Recovery

6.4.1. NMR characterization of cylinder radii distribution using a SHORE-based regularization method.

Participants: Gonzalo Sanguinetti, Matt Hall [Centre for Medical Image Computing, Dept. Computer Science, UCL], Daniel Alexander [Centre for Medical Image Computing, Dept. Computer Science, UCL], Rachid Deriche.

In this work, we are interested in retrieving information about the axon diameter distributions in white matter fiber bundles using NMR, which are commonly modelled as ensembles of parallel cylinders. We add regularization to the 1D-SHORE basis which results in more stable characterization of diameter distributions. To validate the method, we simulate NMR signals using the open source toolkit CAMINO. The results illustrate the enhanced estimation accuracy given by the regularization and provide an alternative validation of the SHORE based method.

This work has been published in [30]

6.4.2. Mapping Average axon diameters under long diffusion time

Participants: Gonzalo Sanguinetti, Rachid Deriche.

This work proposes an original method to recover axon diameter distribution (ADD) parameters using nuclear magnetic resonance. White matter (WM) is modelled as a bi-compartmental medium composed of an intra axonal space where the diffusion is restricted and an extra axonal space where diffusion is hindered. Under the assumption of long diffusion time, we provide a novel and efficient model for the component of the signal due to the restricted part. This technique might be interpreted as an interesting simplification of the AxCaliber framework, which leads to a simpler model and an extremely faster acquisition protocol. To test and validate our method, we use the open-source toolkit Camino for computing Monte-Carlo simulations of NMR data and model the WM as 3D cubic environments, formed by parallel cylinders with gamma distributed radii. Promising experimental results illustrate the potential of the proposed method.

This work has been published in [29] and [28]

6.4.3. Magnitude and complex based diffusion signal reconstruction

Participants: Marco Pizzolato, Timothe Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

In Diffusion Weighted Magnetic Resonance Imaging (DW-MRI) the modeling of the magnitude signal is complicated by the Rician distribution of the noise. It is well known that when dealing instead with the complex valued signal, the real and imaginary parts are affected by Gaussian distributed noise and their modeling can thus benefit from any estimation technique suitable for this noise distribution. We present a quantitative analysis of the difference between the modeling of the magnitude diffusion signal and the modeling in the complex domain. The noisy complex and magnitude diffusion signals are obtained for a physically realistic scenario in a region close to a restricting boundary. These signals are then fitted with the Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE) bases and the reconstruction performances are quantitatively compared. The noisy magnitude signal is also fitted by taking into account the Rician distribution of the noise via the integration of a Maximum Likelihood Estimator (MLE) in the SHORE. We discuss the performance of the reconstructions as function of the Signal to Noise Ratio (SNR) and the sampling resolution of the diffusion signal. We show that fitting in the complex domain generally allows for quantitatively better signal reconstruction, also with a poor SNR, provided that the sampling resolution of the signal is adequate. This applies also when the reconstruction is compared to the one performed on the magnitude via the MLE.

This work has been published in [27]

6.4.4. Extracting a biomarker for the mean cross-sectional area from the ODF

Participants: Rutger H.j Fick, Gonzalo Sanguinetti, Rachid Deriche.

Finding new biomarkers related to the microstructure of white matter (WM) is an active area of research in the MRI community. As opposed to the usual MRI markers such as fractional anisotropy (FA), these biomarkers provide a closer insight on the tissue structure. We introduce a new microstructure based biomarker that is related to the axon diameter distribution (ADD) and can be obtained with a q-space imaging technique like DSI or MAP . This feature is related with the nature and purpose of WM paths in bothnormal and pathological conditions and is obtained from the Orientation Distribution Function (ODF) as twice its maximum value. We show that this value is related with the mean cross-sectional area (MCSA) of an ensemble of parallel axons. The same geometric feature was proposed as a scalar index of microstructure, but was not related to the ODF. In this work we give the formal relation between this microstructure feature and the ODF, and validate it using state-of-the-art numerical simulations.

This work has been published in [18].

6.4.5. An Analytical 3D Laplacian Regularized SHORE Basis and Its Impact on EAP reconstruction and Microstructure Recovery

Participants: Rutger H.j Fick, Demian Wassermann, Gonzalo Sanguinetti, Rachid Deriche.

In diffusion MRI, the reconstructed Ensemble Average Propagator (EAP) from the diffusion signal provides detailed insights on the diffusion process and the underlying tissue microstructure. Recently, the Simple Harmonic Oscillator based Reconstruction and Estimation (SHORE) basis was proposed as a promising method to reconstruct the EAP. However, the fitting of the basis is sensitive to noise. To solve this we propose to use the Laplacian of the SHORE basis as a natural regularization functional. We provide the derivation of the Laplacian functional and compare its effect on EAP reconstruction with that of separated regularization of the radial and angular parts of the SHORE basis. To find optimal regularization weighting we use generalized cross-validation and validate our method quantitatively on synthetic and qualitatively on human data from the Human Connectome Project. We show that Laplacian regularization provides more accurate estimation of the signal and EAP based microstructural measures.

This work has been published in [19]

6.4.6. *Using 3D-SHORE and MAP-MRI to obtain both Tractography and Microstructural Contrasts from a Clinical DMRI Acquisition*

Participants: Rutger H.j Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

Diffusion MRI (dMRI) is used to characterize the directional- ity and microstructural properties of brain white matter (WM) by measuring the diffusivity of water molecules. In clinical practice the number of dMRI samples that can be obtained is limited, and one often uses short scanning protocols that ac- quire just 32 to 64 different gradient directions using a single gradient strength (b-value). Such 'single shell' scanning protocols restrict one to use methods that have assumptions on the radial decay of the dMRI signal over different b-values, which introduces estimation biases. In this work, we show, that by simply spreading the same number of samples over multiple b-values (i.e. multi-shell) we can accurately estimate both the WM directionality using 3D-SHORE and characterize the radially dependent diffusion microstructure measures using MAP-MRI. We validate our approach by undersampling both noisy synthetic and human brain data of the Human Connectome Project, proving this approach is well-suited for clinical applications.

This work has been submitted to ISBI'2015.

6.4.7. *Laplacian-Regularized MAP-MRI Improving Axonal Caliber Estimation*

Participants: Rutger H.j Fick, Demian Wassermann, Gonzalo Sanguinetti, Rachid Deriche.

In diffusion MRI, the accurate description of the entire diffusion signal from sparse measurements is essential to enable the recovery of microstructural information of the white matter. The recent Mean Apparent Propagator (MAP)-MRI basis is especially well suited for this task, but the basis fitting becomes unreliable in the presence of noise. As a solution we propose a fast and robust analytic Laplacian regularization for MAP-MRI. Using both synthetic diffusion data and human data from the Human Connectome Project we show that (1) MAP-MRI has more accurate microstructure recovery com- pared to classical techniques, (2) regularized MAP-MRI has lower signal fitting errors compared to the unregularized approach and a positivity constraint on the EAP and (3) that our regularization improves axon radius recovery on human data.

This work has been submitted to ISBI'2015.

6.4.8. *A Unifying Framework for Spatial and Temporal Diffusion in Diffusion MRI*

Participants: Rutger H.j Fick, Demian Wassermann, Marco Pizzolato, Rachid Deriche.

We propose a novel framework to simultaneously represent the diffusion-weighted MRI (dMRI) signal over diffusion times, gradient strengths and gradient directions. Current frameworks such as the 3D Simple Harmonic Oscillator Reconstruction and Estimation basis (3D- SHORE) only represent the signal over the spatial domain, leaving the temporal dependency as a fixed parameter. However, microstructure- focused techniques such as Axcaliber and ActiveAx provide evidence of the importance of sampling the dMRI space over diffusion time. Up to now there exists no generalized framework that simultaneously models the dependence of the dMRI signal in space and time. We use a functional basis to fit the 3D+t spatio-temporal dMRI signal, similarly to the 3D-SHORE basis in three dimensional 'q-space'. The lowest order term in this expansion contains an isotropic diffusion tensor that characterizes the Gaussian displacement distribution, multiplied by a negative exponential. We regularize the signal fitting by minimizing the norm of the analytic Laplacian of the basis. The continuous 3D+t signal representation can provide new insights on the anomalous nature of the dMRI signal in human tissues, i.e., when mean-squared molecular displacements varies slower than linearly with the diffusion time. From the fitting one can also estimate the axon radius distribution parameters along any direction using approaches similar to AxCaliber. We validate our technique on synthetic data generated using the theoretical model proposed by Callaghan et al. We show that our method is robust to noise and can accurately describe the restricted spatio-temporal signal decay originating from tissue models such as cylindrical pores. Moreover, we apply our method on real data from an ActiveAx acquisition. Overall our approach allows to represent the complete 3D+t dMRI signal which should prove helpful in understanding normal and pathologic nervous tissue.

This work has been submitted to IPMI'2015.

6.4.9. Fast and Robust EAP reconstruction using a Laplacian Regularized SHORE basis and its Impact on Microstructure Recovery

Participants: Rutger H.j Fick, Demian Wassermann, Emmanuel Caruyer [SBIA, University of Pennsylvania Medical School], Rachid Deriche.

In diffusion MRI, the reconstructed Ensemble Average Propagator (EAP) from the diffusion signal provides detailed insights on the diffusion process and the underlying tissue microstructure. Recently, the 3D Simple Harmonic Oscillator based Reconstruction and Estimation (3D-SHORE) basis was proposed as a promising method to reconstruct the EAP. However, the fitting of the basis is sensitive to noise. To solve this we propose to use the Laplacian of the SHORE basis as a natural regularization functional. We provide the derivation of the Laplacian functional and compare its effect on EAP reconstruction with that of separated regularization of the radial and angular parts of the SHORE basis and imposing positive-definiteness in the estimation of the EAP. We validate our method on phantom data with known ground truth and on human data from the Human Connectome Project. We show that Laplacian regularization of the 3D-SHORE basis provides faster and more accurate estimation of the signal and EAP.

This work has been submitted to NeuroImage.

6.5. Functional and structural models analysis

6.5.1. Analyzing Brain Plasticity in Math Learning Using Automated Dissection and Analysis of White Matter Tracts Through dMRI

Participants: Dietsje Jolles [Stanford Medical School], Demian Wassermann, Ritika Chokhani [Stanford Medical School], Jennifer Richardson [Stanford Medical School], Caitlin Tenison [Stanford Medical School], Roland Bammer [Stanford Medical School], Lynn Fuchs [Vanderbit University], Kaustubh Supekar [Stanford Medical School], Vinod Menon [Stanford Medical School].

In a collaboration with Stanford Medical School, we explored longitudinal changes in white matter connectivity triggered by intensive math learning. Plasticity of white matter tracts is thought to be essential for cognitive development and academic skill acquisition in children. However, a dearth of high-quality diffusion tensor imaging (DTI) data measuring longitudinal changes with learning, as well as methodological difficulties in multi-time point tract identification have limited our ability to investigate plasticity of specific white matter tracts. With this contribution, we examined learning-related changes of white matter tracts innervating inferior parietal, prefrontal and temporal regions following an intense two-month math tutoring program. DTI data were acquired from 18 third grade children, both before and after tutoring. A novel fiber tracking algorithm based on a White Matter Query Language (WMQL) was used to identify three sections of the superior longitudinal fasciculus (SLF) linking frontal and parietal (SLF-FP), parietal and temporal (SLF-PT) and frontal and temporal (SLF-FT) cortices, from which we created child-specific probabilistic maps. The SLF-FP, SLF-FT, and SLF-PT tracts identified with the WMQL method were highly reliable across the two time points and showed close correspondence to tracts previously described in adults. Notably, individual differences in behavioral gains after two months of tutoring were specifically correlated with plasticity in the left SLF-FT tract. Our results extend previous findings of individual differences in white matter integrity, and provide important new insights into white matter plasticity related to math learning in childhood. More generally, our quantitative approach will be useful for future studies examining longitudinal changes in white matter integrity associated with cognitive skill development.

This work has been published in [14].

6.5.2. Quantifying Uncertainty in Diffeomorphic Medical Landmark Registration

Participants: Demian Wassermann, Matt Toew [Harvard Medical School - Brigham and Women's Hospital], Marc Niethammer [University of North Carolina at Chapel Hill], William Wells Iii [Harvard Medical School - Brigham and Women's Hospital, MIT].

In a collaboration with Harvard Medical School, the Brigham and Women's Hospital, MIT and the University of North Carolina at Chapel Hill, we proposed a novel mathematical framework to represent uncertainty in diffeomorphic registration techniques. Particularly, we introduced a novel mathematical framework for representing uncertainty in large deformation diffeomorphic image registration. The Bayesian posterior distribution over the deformations aligning a moving and a fixed image is approximated via a variational formulation. A stochastic differential equation (SDE) modeling the deformations as the evolution of a time-varying velocity field leads to a prior density over deformations in the form of a Gaussian process. This permits estimating the full posterior distribution in order to represent uncertainty, in contrast to methods in which the posterior is approximated via Monte Carlo sampling or maximized in maximum a-posteriori (MAP) estimation. The framework was demonstrated in the case of landmark-based image registration, including simulated data and annotated pre and intra-operative 3D images. This type of registration can be extended to several anatomical objects such as white matter tracts represented as streamlines.

This work has been published in [32].

6.5.3. Group Comparisons on White Matter Tracts in Native Space

Participants: Eleftherios Garyfallidis [University of Sherbrooke], Demian Wassermann, Maxime Descoteaux [University of Sherbrooke].

Let us suppose that we want to study specific fiber bundles in different subjects. The common approach would be to use a voxel-wise analyses which will warp scalar volumes in a common space, e.g. MNI space, and show how every subject differentiates from an average template. However, we know that with averaging and warping much of the specific information about the individual subjects' differences is lost. In this work, we provide a solution to this problem by using local streamline registration of specific bundles from different subjects. We show that with this new method we can keep track of the differences from every subject to every other subject in our group study.

This study was performed in collaboration with the SCIL lab of Sherbrook University within the framework of the Brain Connectivities Associate Team and published in [21].

6.5.4. Perfusion Deconvolution via SHORE and Laplacian Regularization

Participants: Marco Pizzolato, Auro Ghosh, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Perfusion imaging comprehensively refers to the recovery of parameters of interest which are related to the passage of blood in the parenchyma (i.e. the functional part) of a tissue. The amount of perfusion is related to both the functionality of the parenchyma and its level of activity. By means of imaging techniques such as Dynamic Susceptibility Contrast MRI it is possible, in each voxel, to measure the tissue concentration $Ct(t)$ of a tracer injected before the scanning in the vascular system. According to the indicator dilution theory1 this is related to the concentration measured in an arterial region $Ca(t)$ described by a convolution with $R(t)$ that is the unknown residue function expressing the remaining time-dependent tracer quantity in the voxel. Historically $R(t)$ is obtained exploiting the convolution theorem $R(t) = FT-1FT[Ct(t)]/FT[Ca(t)]$. However deconvolution is an ill-posed problem making this method very sensitive to noise. Many regularization techniques have been proposed but among all the most adopted technique is truncated Singular Value Decomposition, tSVD. However tSVD is known to underestimate an important perfusion parameter that is the blood flow BF, which can be computed as the maximum peak of the recovered $R(t)$. In this work we propose to use the Simple Harmonic Reconstruction and Estimation framework (SHORE) to estimate $R(t)$ in order to obtain a better parameter estimation. We regularize SHORE using Laplacian regularization. We compare the results with tSVD.

This work has been submitted to ISMRM 2015.

6.5.5. Perfusion MRI Deconvolution with Delay Estimation and Non-negativity Constraints

Participants: Marco Pizzolato, Auro Ghosh, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Perfusion MRI deconvolution aims to recover the time-dependent residual amount of indicator (residue function) from the measured arterial and tissue concentration time-curves. The deconvolution is complicated by the presence of a time lag between the measured concentrations. Moreover the residue function must be non-negative and its shape may become non-monotonic due to dispersion phenomena. We introduce Modified Exponential Bases (MEB) to perform deconvolution. The MEB generalizes the previously proposed exponential approximation (EA) by taking into account the time lag and introducing non-negativity constraints for the recovered residue function also in the case of non-monotonic dispersed shapes, thus overcoming the limitation due to the non-increasing assumtion of the EA. The deconvolution problem is solved linearly. Quantitative comparisons with the widespread block-circulant Singular Value Decomposition show favorable results in recovering the residue function.

This work has been submitted to ISBI 2015.

6.6. Forward and Inverse Problems in MEEG

6.6.1. FindSource3D - Source Localization Using Rational Approximation on Plane Sections

Participants: Todor Jordanov [BESA GmbH, Germany], Jean-Paul Marmorat [École des Mines ParisTech, Sophia Antipolis], Maureen Clerc, Juliette Leblond, Andre Waelkens [BESA GmbH, Germany], Théodore Papadopoulo.

A new method for EEG source localization based on rational approximation techniques in the complex plane was suggested. The method is used in the context of a nested sphere head model, in combination with a cortical mapping procedure [51]. This method was shown to perform perfectly for numerical simulations without noise but its performance with respect to different signal-to-noise ratios (SNRs), to different number of sources and to real EEG data was not investigated until now. The method, formally called FindSource3D (FS3D), is evaluated with data simulations and a real EEG data set.

This work has been published in [40].

6.6.2. Diffusion Magnetic Resonance information as a regularization term for MEG/EEG inverse problem

Participants: Brahim Belaoucha, Anne-Charlotte Philippe, Maureen Clerc, Théodore Papadopoulo.

Several regularization terms are used to constrain the Magnetoencephalography (MEG) and the Electroencephalography (EEG) inverse problem. It has been shown that the brain can be divided into several regions with functional homogeneity inside each one of them. To locate these regions, we use the structural information coming from the diffusion Magnetic Resonance (dMRI) and more specifically, the anatomical connectivity of the distributed sources computed from dMRI. To investigate the importance of the dMRI in the source reconstruction, this work compares the solutions based on dMRI-based parcellation to random parcellation.

This work has been published in [37].

6.6.3. Dictionary learning for multitrial datasets

Participants: Maureen Clerc, Sebastian Hitziger, Théodore Papadopoulo.

Following the path opened with the Consensus matching Pursuit method (CMP) [48], we continue our endeavour to avoid signal averaging using directly the raw signal with the assumption that events of interest are those that repeat in each trial. Towards such a goal, and to improve the simple dictionary used in CMP, we have adapted dictionary learning methods to multitrial bio-electric signals, by explicitly implementing jitter invariance [62]. This allows for a much more detailed data-driven description of events. For example, using local field potential signals of chemically induced spikes (in a rat model), we have been able to distinguish several spike shapes which show some coherence in time. The method has been recently extended to detect spike events in continuous signals (i.e. not organized in epochs). While it requires a good signal to noise ratio, the method is very general and has also been used for various other signal types (see section 6.7).

This work has been published in [39].

6.7. Coupling functional and structural models

6.7.1. Propagation of epileptic spikes revealed by diffusion-based constrained MEG source reconstruction

Participants: Anne-Charlotte Philippe, Théodore Papadopoulo, Christian Bénar [Hospital "La Timone", Marseille], Jean-Michel Badier [Hospital "La Timone", Marseille], Maureen Clerc, Rachid Deriche.

In this work, we study the propagation of an epileptic spike (from single event data). As in the two previous sections, a cortex parcellation is performed using structural information coming from diffusion MRI Then, a MEG inverse problem is defined on a parcellated source space which imposes constant activity on each parcel. This inverse problem is applied separately for measurements obtained in a given time range. The most active parcels over the time range are located and their time course are displayed. This allowed the study of the propagation of an epileptic spike via those active parcels. Results on real data shows varying spatial propagations of an epileptic spike for the same subject.

This work has been published in [41].

6.7.2. Using diffusion MRI information in the Maximum Entropy on Mean framework to solve MEG/EEG inverse problem

Participants: Brahim Belaoucha, Jean-Marc Lina [Centre de Recherches Mathématique, Montréal], Maureen Clerc, Anne-Charlotte Philippe, Christophe Grova [McGill University], Théodore Papadopoulo.

Magnetoencephalography (MEG) and Electroencephalography (EEG) inverse problem is well-known to require regularization in order to avoid ill-posedness. Usually, regularization is based on mathematical criteria (minimum norm, ...). Physiologically, the brain is organized in functional parcels and imposing a certain homogeneity of the activity within these parcels was proven to be an efficient way to analyze the MEG/EEG data. The parcels information can be computed from diffusion Magnetic Resonances Imaging (dMRI) by grouping together source positions shared the same connectivity profile (computed as tractograms from diffusion images). In this work, three parcel-based inverse problem approaches have been tested. The first two approaches are based on minimum norm with added regularization terms to account for the parcel information. They differ by the use of a hard/soft constraint in the way they impose that the activity is constant within each parcel [74]. The third approach is based on the Maximum Entropy on Mean (MEM) framework [42]. It models source activity with a random variable and parcels are also used as a regularization. Several tests have been conducted with synthetic and real data that encompass the MEG/EEG and the diffusion magnetic resonance signals to compare these three approaches in terms of active region-detection accuracy.

This work has been published in [36].

6.8. Brain Computer Interfaces

6.8.1. CoAdapt P300 speller: optimized flashing sequences and online learning

Participants: Maureen Clerc, Théodore Papadopoulo, Loïc Mahé.

Our work in Brain Computer Interfaces was centered around the visual P300 speller system: a virtual keyboard allowing to type words by detecting the P300 wave, an automatic deflection of the central component of the electric potential, occurring approximately 300 ms after the apparition of an intermittent and rare event, on which the user's attention is focussing. The idea behind the P300 speller is very simple: the system displays series of stimuli (flashes), over the keyboard elements, and detects whether or not the EEG recorded after each flash contains a P300. Its implementation is not so simple, because of the low amplitude of the P300 compared to the background EEG, and of the inter-subject variability of this signal.

The advantage of this system is not to require any training on the part of the user. However the BCI system has to be trained to detect the P300 component from the background EEG: this is done through a calibration phase.

We developped a new method to reduce the calibration phase, with a transfer learning method called "mixture of experts" (MOE). The MOE classifier makes its predictions by averaging the decisions from a pre-recorded database of classifiers coming from other recording sessions (with other subjects) [31]. The decisions were made by using an evidence accumulation scheme, which updated the prediction at every flash of the keyboard [17].

Part of this work has been implemented in the software: CoAdapt P300 Stimulator.

6.8.2. P300 speller: clinical feasibility study with Amyotrophic Lateral Sclerosis

Participants: Maureen Clerc, Théodore Papadopoulo, Loïc Mahé, Asya Metelkina, Violaine Guy [Nice University Hospital], Claude Desnuelle [Nice University Hospital].

From September 2013 to July 2014, we were very involved in running an experiment with the Centre de Référence Maladies Neuromusculaires et SLA (CRMN/SLA) of Nice University Hospital. This study, partly funded by "Assocation pour la Recherche sur la Sclérose Latérale Amyotrophique", was conducted on 20 patients, who routinely come to be examined at the hospital. Each patient came for 3 sessions where he/she was allowed to use the P300 speller, after being equipped with an electro-encephalography cap, and watching a video explaining the modus operandi of the P300 and on their role in the study.

The P300 speller system has been organized in a way to make it relatively easy to deploy in a clinical setting: it involves only one laptop, and requires limited intervention from the caregiver. The most intricate operation is to position the EEG headset and ensure a correct impedance (below 5 kΩ) for all electrodes.

Each session consisted of three blocks, after the initial calibration phase: a copy spelling task of two ten-letter words, a free spelling task of approximately twenty characters, and an optional block of free use of the system for writing. Finally, the patient was asked to answer a questionnaire. This study intends to investigate the feasibility of setting up and using the P300 speller, from an operational point of view at the hospital. Translational studies of this type are extremely important for the adaptation the BCI systems to the target patient populations, and a large-scale usability study for the P300 speller has never been done before in France.

6.8.3. BCI Challenge: A spell on you if you cannot detect errors!

Participants: Maureen Clerc, Théodore Papadopoulo, Jérémie Mattout [Centre de Recherche en Neurosciences de Lyon, INSERM], Emmanuel Maby [Centre de Recherche en Neurosciences de Lyon, INSERM].

We have proposed an international BCI Challenge on decoding Error Potential signals. The winners will be announced at the 7th International IEEE/EMBS Conference on Neural Engineering, Montpellier, in April 2015. The Challenge was open on the Kaggle platform on Nov 14, 2014 and will close on Feb 24, 2015, see: website. At this date the competition is still open, and it has so far attracted 212 participants forming 181 teams.

In the P300 speller paradigm (see above) and in other BCI where a discrete feedback can be presented to the user, the EEG evoked response to the feedback can be recorded and processed online in order to evaluate whether the item selection was correct or not. This decision, if reliable, could then be used to improve the BCI performance by implementing some error correction strategy. In this competition, participants are asked to submit an Error Potential detection algorithm, capable of detecting the erroneous feedbacks online and to generalize across subjects (transfer learning).

The data used in this competition was acquired in the scope of the CoAdapt ANR project.

DEMAR Project-Team

5. New Results

5.1. Modelling and identification of the sensory-motor system

5.1.1. Whole Body Center of Mass Estimation with Portable Sensors: Using the Statically Equivalent Serial Chain and a Kinect

Participants: Alejandro Gonzalez de Alba, Mitsuhiro Hayashibe, Vincent Bonnet, Philippe Fraisse.

The trajectory of the whole body center of mass (CoM) is useful as a reliable metric of postural stability. If the evaluation of a subject-specific CoM were available outside of the laboratory environment, it would improve the assessment of the effects of physical rehabilitation. A method is developed tot enable tracking CoM position using low-cost sensors such that it can be moved around by a therapist or easily installed inside a patientâs home. We compare the accuracy of a personalized CoM estimation using the statically equivalent serial chain (SESC) method and measurements obtained with the Kinect to the case of a SESC obtained with high-end equipment (Vicon). We also compare these estimates to literature-based ones for both sensors. The method was validated with seven able-bodied volunteers for whom the SESC was identified using 40 static postures. The literature-based estimation with Vicon measurements had a average error 24.9 ± 3.7 mm; this error was reduced to 12.8 ± 9.1 mm with the SESC identification. When using Kinect measurements, the literature-based estimate had an error of 118.4 ± 50.0 mm, while the SESC error was 26.6 ± 6.0 mm. The subject-specific SESC estimate using low-cost sensors has an equivalent performance as the literature-based one with high-end sensors. The SESC method can improve CoM estimation of elderly and neurologically impaired subjects by considering variations in their mass distribution.

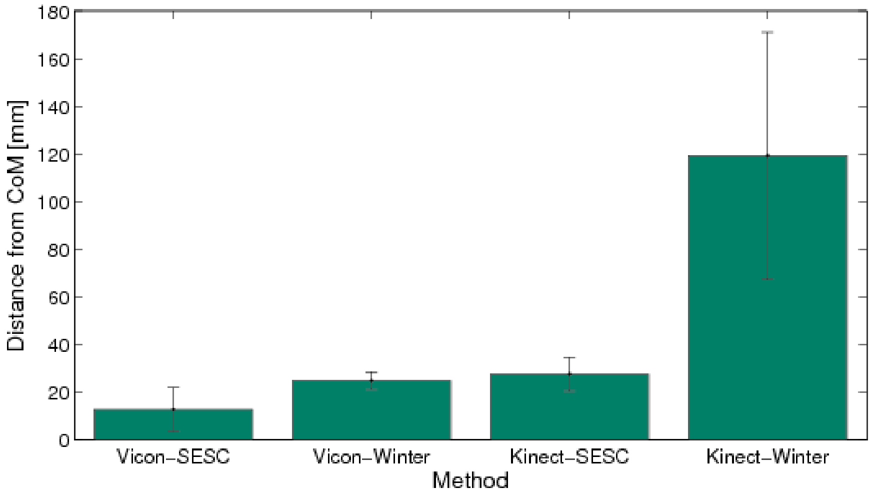


Figure 2. The bars correspond to the averaged results for all seven subjects. We observe an increase in the accuracy of the identified SESCs with respect to the literature estimates; we found that the performance of the Kinect-SESC is equivalent to that of the literature-based estimate using high-end sensors.

5.1.2. A Personalized Balance Measurement for Home-based Rehabilitation

Participants: Alejandro Gonzalez de Alba, Philippe Fraisse, Mitsuhiro Hayashibe.

We use the subject-specific center of mass (CoM) estimate offered by the statically equivalent serial chain (SESC) method and the zero rate of change of angular momentum (ZRAM) concept to evaluate balance for a series of dynamic motions. Two healthy subjects were asked to stand on a Wii balance board and their SESC parameters were identified. A set of dynamic motions was to evaluate the rate of change of centroidal angular momentum and the distance of the ZRAM point to the center line of the support polygon. We found a good match between both balance metrics. As an application example, the subjects performed a dynamic motion (such as walking and abruptly stopping) and the stability was determined in real-time using the ZRAM point from the personalized CoM trajectory. This was implemented with a real-time balance visualization tool based on Kinect measurements for home-based rehabilitation.

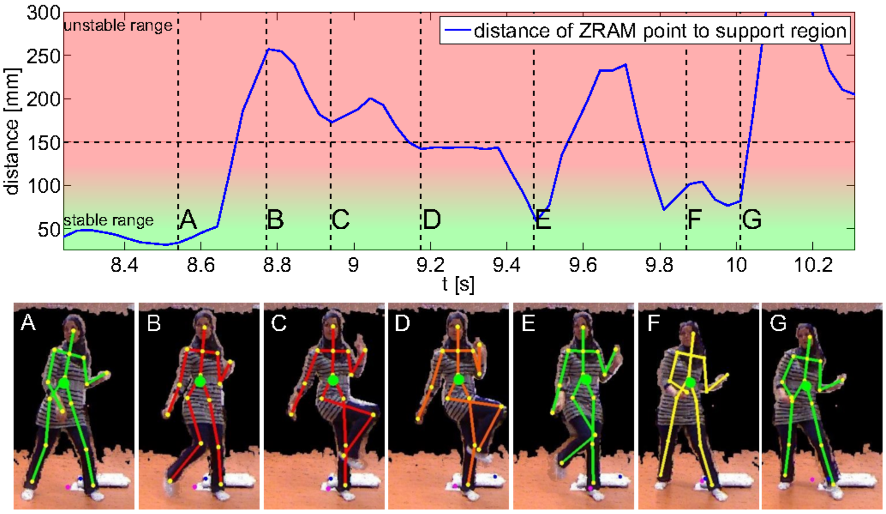


Figure 3. The distance between the ZRAM point and the center of the support polygon can be used to determine balance in real-time and shown as feedback during a physical rehabilitation program. In this example, the subject's skeleton is colored in red to indicate an unstable movement and in green for a stable one.

5.1.3. Methodology for automatic movement cycle extraction using Switching Linear Dynamic System

Participants: Roberto de Souza Baptista, Mitsuhiro Hayashibe, Antônio P. L. Bó [Univ. Brasilia].

Human motion assessment is key for motor-control rehabilitation. Using standardized definifions and spatiotemporal features - usually presented as a movement cycle diagram- specialists can associate kinematic measures to progress in rehabilitation therapy or motor impairment due to trauma or disease. Although devices for capturing human motion today are cheap and widespread, the automatic interpretation of kinematic data for rehabilitation is still poor in terms of quantitative performance evaluation. In this paper we present an automatic approach to extract spatiotemporal features from kinematic data and present it as a cycle diagram. This is done by translating standard definitions from human movement analisys into mathematical elements of a Switching Linear Dynamic System model. The result is a straight-forward procedure to learn a tracking model from a sample execution. This model is robust when used to automatically extract the movement cycle diagram of the same motion (the Sit-Stand-Sit, as an example) executed in different subject-specific manner such as his own motion speed.

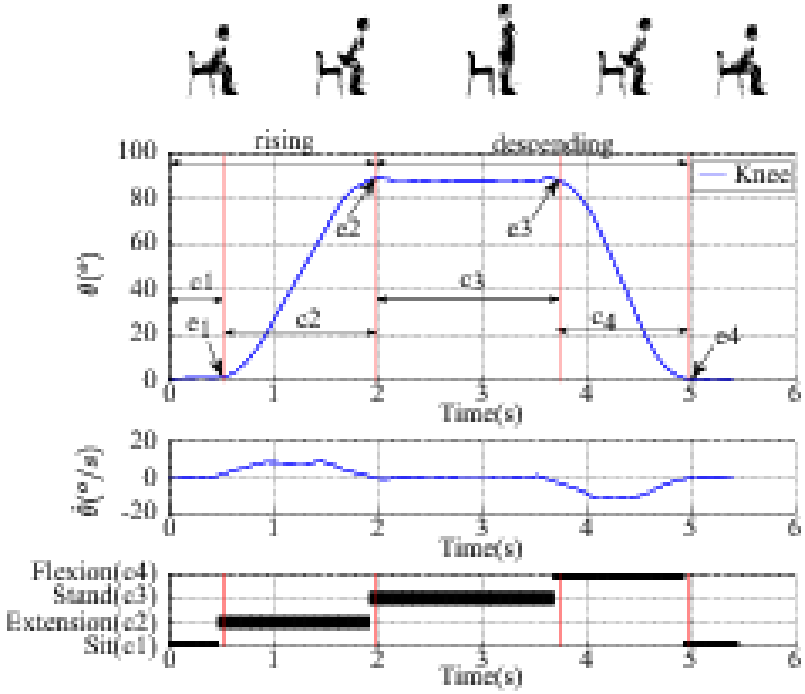


Figure 4. Training dataset consisting of one execution of the Sit-Stand-Sit movement cycle. Events (e_i), components (c_i) and the rising and descending phases are marked. θ and $\dot{\theta}$ indicates angle and angular velocity.

5.1.4. Real-time Muscle Deformation via Decoupled Modeling of Solid and Muscle Fiber Mechanics

Participants: Yacine Berranen, Mitsuhiro Hayashibe, David Guiraud, Benjamin Gilles.

This work presents a novel approach for simulating 3D muscle deformations with complex architectures. The approach consists in choosing the best model formulation in terms of computation cost and accuracy, that mixes a volumetric tissue model based on finite element method (3D FEM), a muscle fiber model (Hill contractile 1D element) and a membrane model accounting for aponeurosis tissue (2D FEM). The separate models are mechanically binded using barycentric embeddings. Our approach allows the computation of several fiber directions in one coarse finite element, and thus, strongly decreases the required finite element resolution to predict muscle deformation during contraction. Using surface registration, fibers tracks of specific architecture can be transferred from a template to subject morphology, and then simulated. As a case study, three different architectures are simulated and compared to their equivalent one dimensional Hill wire model simulations.

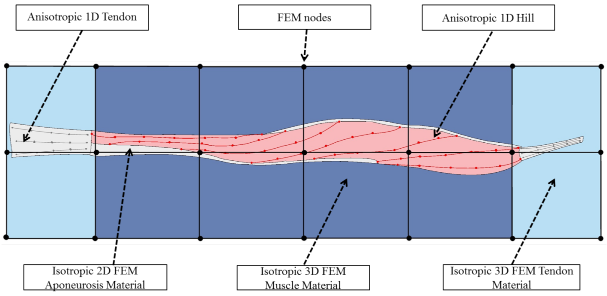


Figure 5. Muscle multi-model scheme: The different models are linked via barycentric embeddings. This approach strongly decreases the required finite element resolution to predict muscle deformation during contraction

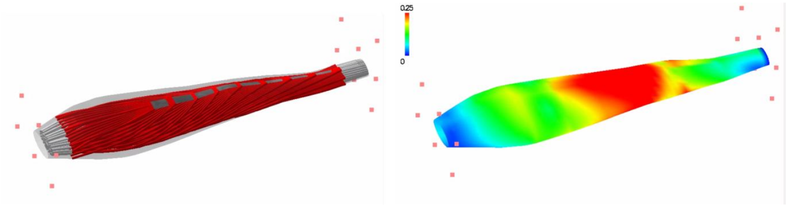


Figure 6. Left: Hybrid model of bipennate muscle, right: internal stress due to isometric contraction

5.1.5. Adaptive model for viscoelastic solids

Participants: Benjamin Gilles, Maxime Tournier, Matthieu Nesme, Francois Faure.

A new adaptive model for viscoelastic solids is presented in [50], [34]. Unlike previous approaches, it allows seamless transitions, and simplifications in deformed states. The deformation field is generated by a set of physically animated frames. Starting from a fine set of frames and mechanical energy integration points, the model can be coarsened by attaching frames to others, and merging integration points. Since frames can be attached in arbitrary relative positions, simplifications can occur seamlessly in deformed states, without returning to the original shape, which can be recovered later after refinement. We propose a new class of velocity-based simplification criterion based on relative velocities. Integration points can be merged to reduce the computation time even more, and we show how to maintain constant elastic forces through the levels of detail. This meshless adaptivity allows significant improvements of computation time.

5.1.6. Functional Brain Stimulation: Filling the gap between micro- and direct electrical stimulation of the brain in order to better understand and innovate

Participants: Marion Vincent, Olivier Rossel, Mitsuhiro Hayashibe, Guillaume Herbet, Hugues Duffau [Neurosurgery Department, CHU, Montpellier], David Guiraud, François Bonnetblanc.

Micro-stimulation (MES) and Direct electrical stimulation (DES) of the brain are both used to perform in vivo functional mapping of the brain in fundamental neuroscience and neurosurgery respectively. The former is performed in animal experiments while the latter is performed on humans in the operative room. Very recently, a strong debate occurred to determine whether DES used during “wide-awake surgery” with success is a gold standard to study brain functions (Mandonnet et al., 2010; Borchers et al. 2012; Desmurget et al., 2013). In this debate, confusion is very often made between DES and MES, as these are considered to induce similar effects on the nervous tissues, with comparable behavioural consequences. However, electrical stimulation (ES) parameters used in both techniques are clearly different. More surprisingly, a strong biophysical rational of their choices is lacking. It may be due to historical, methodological and technical constraints that have guided empirically them. These constraints may have strongly shaped and limited experimental protocols in a standard way. By contrast, the gap between MES and DES may reveal a great potential for new experimental paradigms in ES of the brain in vivo. By considering this gap and new technical developments in the design of stimulators, it may be time to move on to alternative and innovative stimulations protocols, especially regarding and inspired from what is performed in functional electrical stimulation (FES) of peripheral nerves, for which more theoretical supports exist.

5.1.7. Modelling of structural and functional brain connectivity networks in Diffuse Low-Grade Glioma (DLGG)

Participants: Jija Syamala James, Anirban Dutta, François Bonnetblanc, David Guiraud, Nicolas Menjot de Champfleur [Department of Neuroradiology,CHU, Montpellier], Emmanuelle Le Bars [Institute of Human Functional Imaging 12 FH, CHU, Montpellier], Hugues Duffau [Neurosurgery Department, CHU, Montpellier].

The impairment of functional brain connectivity networks in Diffuse Low Grade Glioma (DLGG) subjects can lead to distinct functional deficits where the challenge remains in greater understanding of distribution of dynamic brain connectivity. Our multimodality (resting state functional MRI, Diffusion Tensor Imaging and tractography) neuroimaging study aims to evaluate the differences in spatial and temporal patterns of brain connectivity networks in DLGG patients, pre and postoperatively.

In order to identify the brain connectivity networks, we analysed resting state fMRI data of 22 mesio-frontal DLGG patients by using MELODIC (Multivariate Exploratory Linear Optimised Decomposition into Independent Components)-ICA module, implemented in Freesurfer Software Library [FSL] (www.fmrib.ox.ac.uk/fsl). First we evaluated the clinical efficacy of rsfMRI technique to non-invasively map the dynamic functional reorganizations of brain connectivity networks in glioma subjects. Further we observed the existence of altered inter-hemispheric functional connectivity in DLGG subjects postoperatively.

We are currently performing this analysis using larger sample size in order to find out

a) the invivo structure-function relationship of motor, language and visual brain networks by fusing the information from DTI and fMRI.
b) the correlation between brain connectivity networks and neurobehavioural performance.
c) the neuroplastic alterations in topographic organization and strength of connections before and after DLGG surgery.

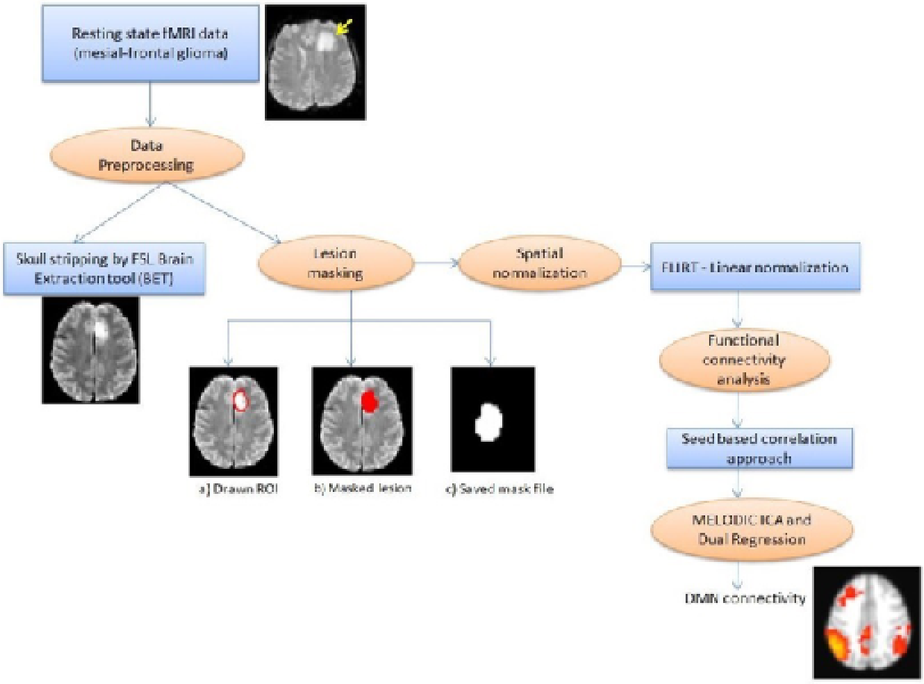


Figure 7. Steps involved in resting state functional connectivity analysis.

5.1.8. SPINSTIM: Direct spinal stimulation for rehabilitation of bladder, bowel and sexual functions in spinal cord injury

Participants: Christine Azevedo Coste, Luc Bauchet [Neurosurgery Department, CHU, Montpellier], Claire Delleci [CHU Bordeaux], Charles Fattal, Thomas Guiho, David Guiraud, Jean-Rodolphe Vignes [CHU Bordeaux].

For the general public, spinal cord injury (SCI) is often restricted to limb paralysis. In reality, by interrupting communication between encephalon and peripheral organs, medullary wounds lead to physiological deficiencies such as urinary (retention and/or incontinence), gastrointestinal (constipation) or sexual impairments; disorders which are the center of patient's expectations.

Spinal cord stimulation (SCS) is a general term which includes both epidural and intradural stimulation. Originally associated with the treatment of chronic neurogical pain (in the 1970ies), SCS led also to immediate and profound improvements of sensory and motor functions in recent studies both on SCI patients (few subjects involved) and rodents.

Despite these promising results some limitations have still to be overcome. Among them, the use of small animal models, the empirical aspect of the stimulation procedure and the impact of these protocols on intestinal and urinary functions are critical.

To counteract these limits, we want to explore intradural and epidural stimulations in an intermediate model- the house pig- and assess their impact on bladder, guts and genitals. In order to evaluate our approach, we will record EMG signals of lower limbs and sphincters (both urethral and anal), and simultaneously, we will monitor bladder and rectal pressure.

5.1.8.1. Development of point of care testing (POCT) device for neurovascular coupling from simultaneous recording of EEG and NIRS during non-invasive brain stimulation (NIBS) for closed-loop control of NIBS.

Participants: Mehak Sood, Utkarsh Jindal, Mitsuhiro Hayashibe, Stephane Perrey, Michael A. Nitsche, Abhijit Das, Shubhajit Roy Chowdhury, Anirban Dutta.

Our preliminary work showed that transcranial direct current stimulation (tDCS) can perturb local neuronal activity which can be used for assessing regional neurovascular coupling (NVC) functionality. It was postulated that tDCS leads to rapid dynamic variations of the brain cell microenvironment that perturbs the hemodynamic and electromagnetic responses. Based on these preliminary studies, we developed a POCT device for EEG-NIRS based screening and monitoring of neurovascular coupling functionality under perturbation with tDCS. The stroke case study showed detectable changes in the degree of NVC to a 0.526A/m^2 square-pulse (0-30sec) of anodal tDCS where these alterations in the vascular system may result in secondary changes in the cortical excitability. The objective of this case study was to evaluate an empirical method to assess NVC using cross-correlation function (CCF) between mean (cortical) tissue oxy-haemoglobin concentration time-series and averaged PSD time-course from the EEG spectrogram. The CCF based assessment of the patient-specific status of NVC are currently being studied in a larger cohort with small vessel diseases. The overarching goal is closed-loop control of tDCS based on simultaneous recording of EEG and NIRS during non-invasive brain stimulation.

5.2. Synthesis and Control of Human Functions

5.2.1. Ergonomics of the control by a quadriplegic of hand functions

Participants: Christine Azevedo Coste, David Guiraud, Wafa Tigra, Charles Fattal.

In subjects with complete Spinal Cord Injury (SCI) above C7, the four limbs are paralyzed (quadriplegia). Recovery of grasping movements is then reported as a priority. Indeed, most activities of daily living are achieved through upper limbs. Thus, restoration of hand and forearm active mobility could significantly increase independence and quality of life of these people and decrease their need of human aid. Although most of the subjects plebiscite pharmacological or biological solutions, only orthotics and Functional Electrical Stimulation (FES) allow, so far, to restore hand movements but they are rarely used. Ergonomics and comfort

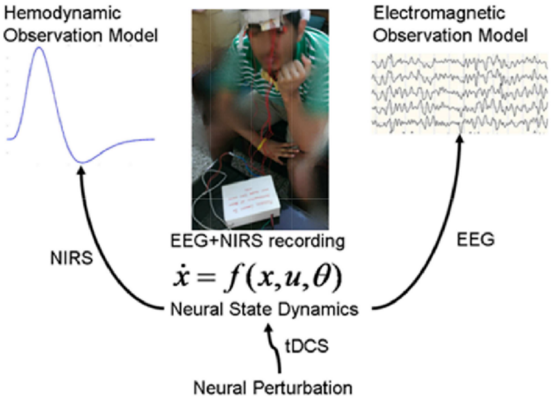


Figure 8. EEG-NIRS based simultaneous recording of the hemodynamic and electromagnetic responses to perturbations with transcranial direct current stimulation (tDCS).

of piloting mode could partly explain the low usage of these systems. In this context, our aim is to explore possible solutions for subjects to interact with such devices. In this article we propose to evaluate the capacity of active upper limb muscles contraction to be used to intuitively control FES in tetraplegic subjects. In this study, we assessed the ability to gradually contract different muscles: trapezius, deltoid, platysma and biceps. Three subjects with C6 to C7 neurological level of lesion were included. We show that over the active upper limb muscles tested, contraction of the trapezius muscle was considered by the subjects as the most comfortable and could be employed as an intuitive mode of control of functional assistive devices.

5.2.2. Implementation of filtering, calibration and reconstruction algorithms dedicated to the use of inertial measurement units related to rehabilitation and movement analysis

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet.

This work has been done within SENSBIO ADT.

- Stride length estimation

  An algorithm has been implemented for the sep by step stride length calculation from only one shank located sensor algorithm (calibration, segmentation and reconstruction) Experimental validation was done on 10 healthy subjects (error <10%), 12 Parkinson disease subjects, and 7 hemiparetic subjects.
- Comparison of different reconstruction algorithms

  A Python programmed toolbox was developed for movement analysis (SensbioTK) in order to compare the performances of Mahony, Madgwick, Martin-Salaun sensor fusion algorithms.
- Real-time MoCap data processing Matlab based software

  The algorithm is based on Nexus SDK for rotation angle and translation calculation of a tracked object.

5.2.3. Dominant Component in Muscle Fatigue Induced Hand Tremor during Laparoscopic Surgical Manipulation

Participants: Sourav Chandra, Mitsuhiro Hayashibe, Asokan Thondiyath [IIT Madras].

Accuracy of laparoscopic surgery gets affected by the hand tremor of the surgeons. Though cognitive load is inevitable in such activity which promotes tremor, muscle fatigue induced tremor is significant among the most important source of tremor. Characteristic of fatigue induced hand tremor and its dominant directional properties are reported in this work. For a fixed laparoscopic tool grip with temporally synchronized predefined task protocols, characteristics of fatigue induced tremors have been studied. Dominant component of tremor is found to be in the sagittal plane in case of both static and dynamic tasks. The results shown in the figure, sagittal plane (z axis) component of hand tremor is higher than the other directions. In order to relate it with the muscle fatigue level, spectral properties of surface electromyography (SEMG) were also investigated simultaneously. A study of transient effect on tool positioning was also included, which conjointly advocates the other experimental results on fatigue induced hand tremor as well. Currently a better metric for muscle fatigue is being analyzed and studied with a purview of relation in between SEMG and hand tremor (fig.9).

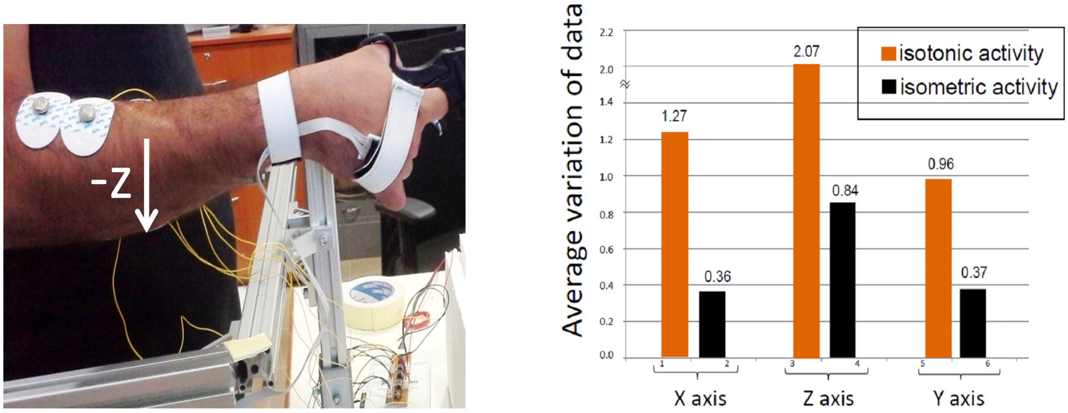


Figure 9. (left) Experimental hand tremor measurement setup (right) Dominant component of hand tremor in all three planes

5.2.4. Human arm optimal motion analysis in industrial task for improving ergonomics of painful workstations

Participants: Nahema Sylla, Vincent Bonnet [TUAT, Japan], Nahid Armande [PSA], Philippe Fraisse.

In PSA Peugeot Citroen factories, high precision requirements of workstations make them being manual. One of the main goal of the car manufacturer is to minimize the pain of workers while maintaining high efficiency of production lines. Consequently, assisting operators with an exoskeleton is a potential solution for improving ergonomics of painful workstations while respecting industrial constraints [33]. We have developped a new approach based on inverse optimization to better understand human arm motion in industrial screwing task. The process combines several criteria to minimize such as energy expenditure or trajectory smoothness leading to the optimal trajectory of a typical screwing task, often performed by workers. Estimated joint trajectories are similar with the measured ones, with a mean square error of 4 degrees. The resulting cost-function is mainly composed of energy expenditure and geodesic criteria. Results show the relevance of using composite cost function in human motion planning [52]. This method has been applied to evaluate a 7 DoF exoskeleton in terms of motion control. The results of our study show that the hybrid composition of the free arm movement was accurately determined. At contrary, when wearing the exoskeleton, which produces an arbitrary determined torque compensation, the motion is different from the naturally adopted one. This study is part of the evaluation and comprehension of the complex neuromuscular mechanism resulting in wearing an exoskeleton several hours per day for industrial tasks assistance [49].

5.2.5. A System for Real-time Online Estimation of Joint Torque with Evoked EMG under Electrical Stimulus

Participants: Zhan Li, Mitsuhiro Hayashibe, David Andreu, David Guiraud.

Functional electrical stimulation (FES) is a useful rehabilitation technique for restoring motor capability of spinal cord injured (SCI) patients by artificially driving muscle contraction through delivering electrical pulses. Real-time FES systems with online modulation ability are in great demand towards clinic applications. In this work, online estimation of joint torque with evoked electromyography (eEMG) in real-time environment is presented. The eEMG is acquired by National Instrument (NI) acquisition card and the stimulus is produced by wireless stimulator (manufactured by Vivaltis Inc., France). Kalman filter (KF) is adopted and embedded as the online estimator. Such real-time online torque estimation system produces promising results. Currently, the implementation to use EMG signal also from wireless Pod module is under investigation (fig.10).

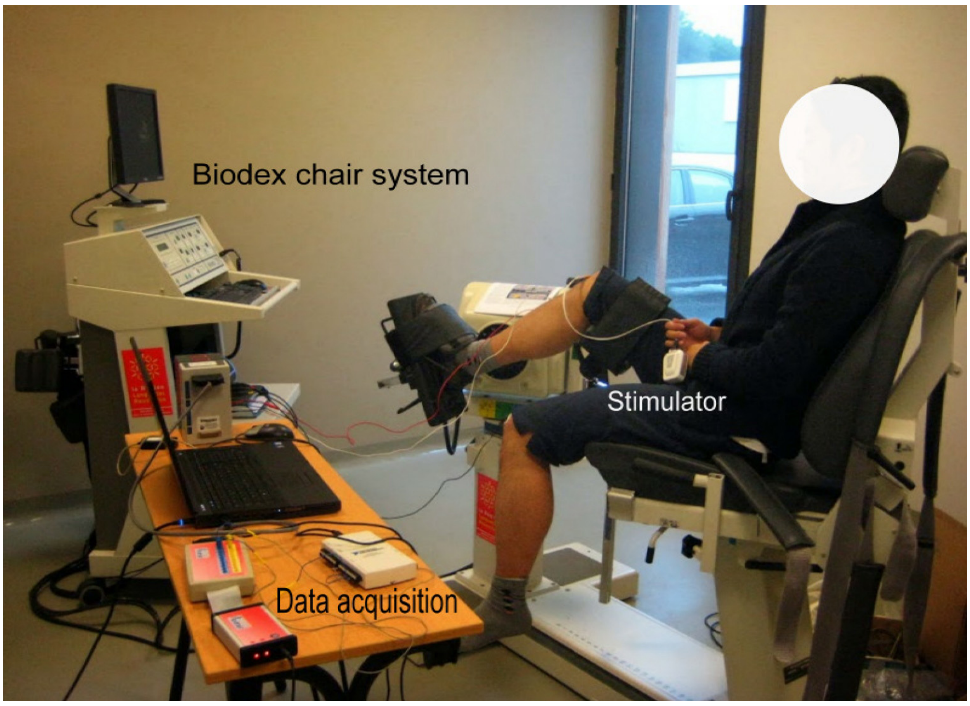


Figure 10. Real-time online FES-induced torque estimation system

5.2.6. Freezing of Gait Analysis and Detection

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet, Christian Geny [CHU Montepllier, Neurologie].

We have extended and optimized the work on Freezing Of Gait (FOG) and destination detection (Maud Pasquier PhD thesis, 2013). A new software for FOG Criteria based on cadency and stride length calculation has been implemented. Results were compared to Freezing Index (Moore et al., 2008) based on frequencies analysis of legs vertical acceleration. The comparison between detection and reality is done on the basis of video analysis of the performed tasks. A software has been implemented for the video labelization Matlab/VLC based tool, with graphical user interface, for marking and synchronizing events from a video (MovieFOG). 14 Parkinson disease patients were included in the study [12][48] (fig.11).

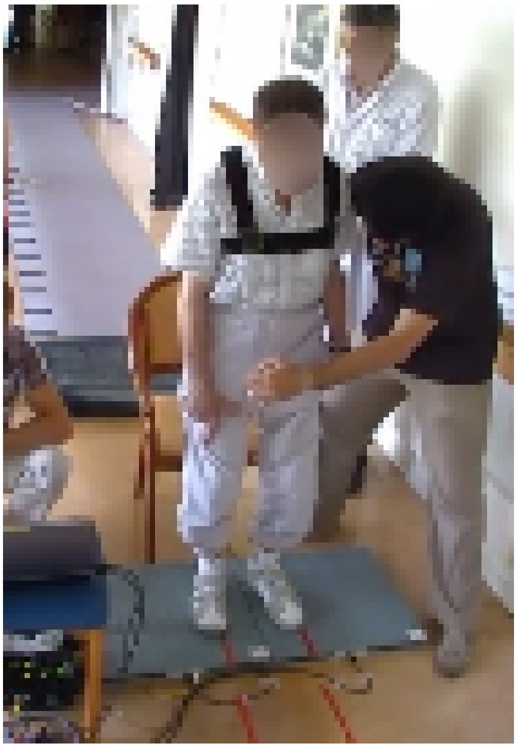


Figure 11. Freezing of Gait Analysis and Detection

5.2.7. A novel brain-computer interfacing paradigm for control of multi-DOFs robot with adaptive EEG decoding and synergetic environment adaptation

Participants: Saugat Bhattacharyya, Shingo Shimoda [RIKEN, Japan], Mitsuhiro Hayashibe.

The study proposes a novel brain-computer interfacing paradigm for control of a multi-joint redundant robot system. Here, the user would determine the direction of end-point movement of a 3-dof robot arm using motor imagery electroencephalography (EEG) signal with co-adaptive decoder while a synergetic motor learning algorithm manages a peripheral redundancy in multi DOF joints toward energy optimality through tacit learning. As in human motor control, torque control paradigm is employed for a robot to be sensitive to the given environment. The dynamic condition of the robot arm is taken into consideration by the learning algorithm. Thus, the user needs to only think about the end-point movement of the robot arm, which allows simultaneous multi-joints control by BCI. The k-Nearest Neighbor based decoder designed for this study is adaptive to the changing mental state of the user. Offline experiments on the decoder reveals that its classification accuracy gradually increases at each learning stage. Online experiments also reveals that the users successfully reach their targets with an average decoder accuracy of more than 65% in different end-point load conditions. The details of the BCI control paradigm, shown in Fig.12 , is as follows: Initially, the robot is trained to its dynamic environment using a tacit learning approach for a fixed period of time. In this study, the load carried by the robot is treated as the environmental changes along with link segment inertial configuration changes. As a result, the movement of the joints of the robot adapts to the changing load. After the training of the robot, the subject begins his/her task of visualizing the target and decides on the direction of motion of the robot. Here, we have used left and right movement imagery to move the robot in upward and downward direction, respectively. Subsequently, the EEG signals (corresponding to the movement imagined) are pre-processed and their corresponding features are extracted. These features are used as inputs to the decoder which determines the mental state of the user and sends command to the robot to move in the equivalent direction.

5.2.8. Impact of the gaze direction on the skier trajectory

Participants: Christine Azevedo Coste, Benoît Sijobert, Roger Pissard-Gibollet, Nicolas Coulmy [FFS annecy].

This work is done within a collaboration with the French Skiing Federation (FFS). Preliminary indoor tests for designing an experimental protocol using motion capture and inertial sensors was realized using a ski simulator and Motion Capture Tool from DEMAR (VICON BONITA). Indoor and outdoor experimentations were done with olympic skiers. In outdoor tests IMUs were combined with SMI eye tracking device. Preliminary analysis and data reconstruction has been done. A VPython script for 3D visualizing of skier movements was developed (fig.13).

5.2.9. Development of a low-cost biofeedback system for electromyogram-triggered functional electrical stimulation therapy in conjunction with non-invasive brain stimulation

Participants: Anirban Dutta, Christine Azevedo Coste, Mitsuhiro Hayashibe, Uttama Lahiri, Abhijit Das, Michael A. Nitsche, David Guiraud.

Functional electrical stimulation (FES) facilitates ambulatory function after paralysis by activating the muscles of the lower extremities. The FES-assisted stepping can either be triggered by a heel-switch, or by an electromyogram-(EMG-) based gait event detector. A group of six chronic (>6 months post-stroke) hemiplegic stroke survivors underwent transcutaneous FES-assisted training for 1 hour on stepping task with EMG biofeedback from paretic tibialis anterior (TA) and medial gastrocnemius (GM) muscles, where the stimulation of the paretic TA or GM was triggered with surface EMG from the same muscle. During the baseline, post-intervention, and 2-day-postintervention assessments, a total of 5 minutes of surface EMG was recorded from paretic GM and TA muscles during volitional treadmill walking. Two-way ANOVA showed significant effects in terms of P-values for the 6 stroke subjects, 0.002, the 3 assessments, 0, and the interaction between subjects and assessments, 6.21E-19. The study showed a significant improvement from baseline in paretic GM and TA muscles coordination during volitional treadmill walking. Moreover, it was found that the EMG-triggered

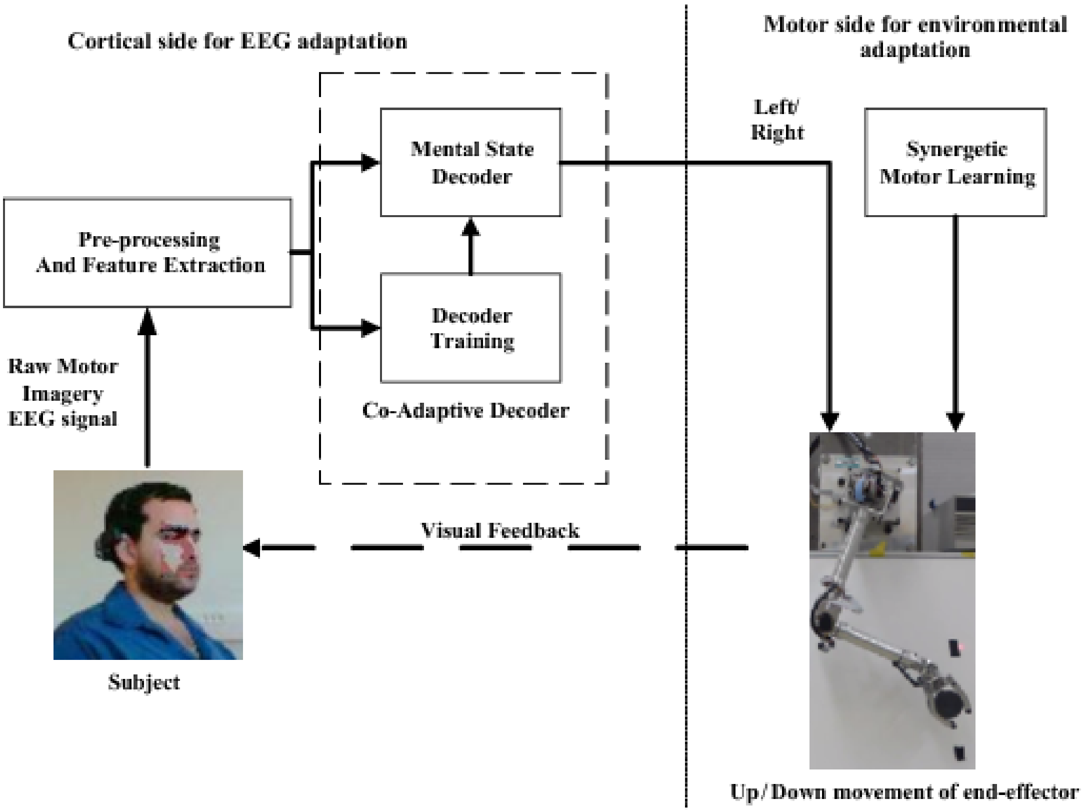


Figure 12. BCI paradigm employed in this study for control of multi-DOFs robot using adaptive left-right motor imagery decoder and synergetic motor learning for peripheric redundancy management (via tacit learning).

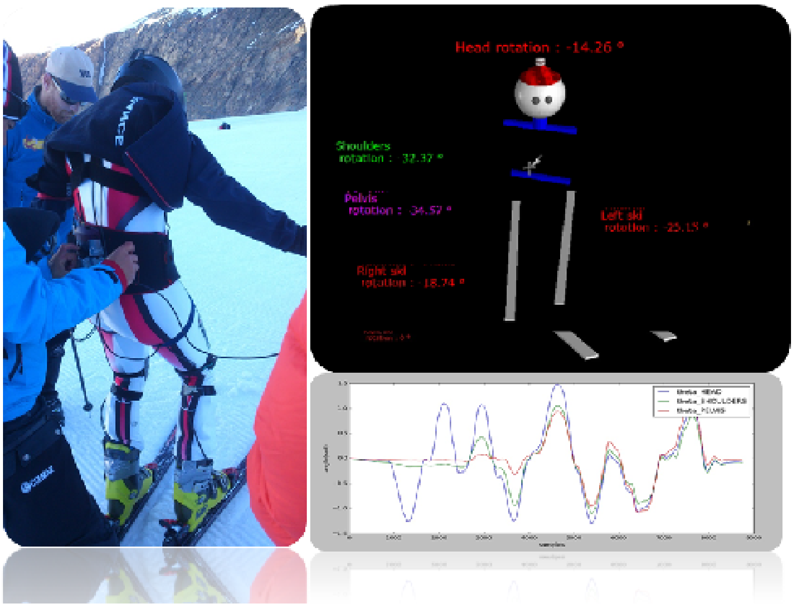


Figure 13. Impact of the gaze direction on the skier trajectory

FES-assisted therapy for stand-to-walk transition helped in convergence of the deviation in centroidal angular momentum from the normative value to a quasi-steady state during the double-support phase of the nonparetic. Also, the observational gait analysis showed improvement in ankle plantarflexion during late stance, knee flexion, and ground clearance of the foot during swing phase of the gait. Currently, we are conducting preliminary stroke studies to evaluate non-invasive brain stimulation as an adjunct to EMG-triggered FES therapy for movement rehabilitation [19][22].

5.2.10. Development and bench-testing of a low-cost eye tracking system (ETS) to measure gaze abnormality in stroke towards virtual reality based visuomotor therapy task.

Participants: Deepesh Kumar, Abhijit Das, David Guiraud, Michael A. Nitsche, Anirban Dutta, Uttama Lahiri.

We conducted a preliminary usability study while incorporating our novel low-cost ETS to measure one's eye gaze indices in response to presented visual task. The ETS provided gaze-related biomarkers which has the potential to be mapped to the probable abnormalities in one's eye movement pattern in stroke. Our preliminary findings with stroke-survivors and age-matched healthy participants indicate the potential of our low-cost ETS to provide quantitative measures of the difference in gaze-related biomarkers between the two groups of participants. Based on these preliminary results, we are conducting a clinical stroke study on ETS based screening and monitoring of performance during a virtual reality based visuomotor balance therapy task.

5.3. Neuroprostheses and technology

5.3.1. Abstraction and composition for formal design of neuroprotheses

Participants: David Andreu, Hélène Leroux, Karen Godary-Dejean.

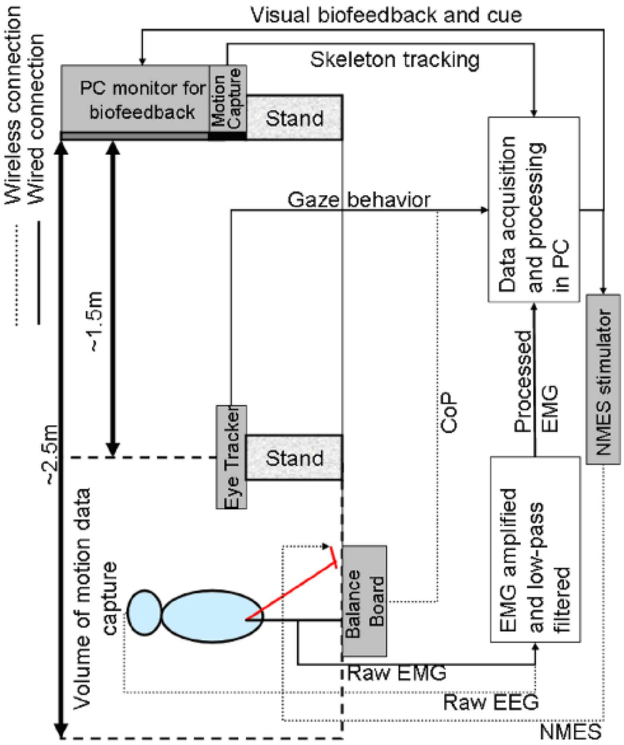


Figure 14. Human-machine-interface integrating low-cost sensors for post-stroke balance rehabilitation

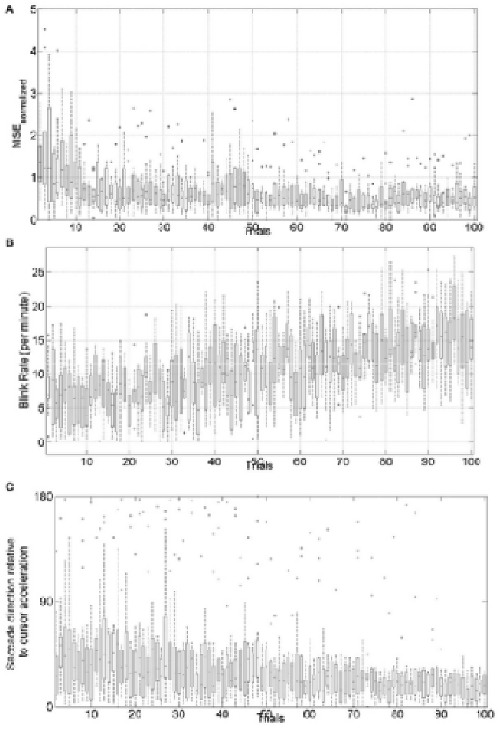


Figure 15. A. Box-plot of normalized mean squared error (MSE) across 10 healthy subjects, B. Box-plot of the blink rate during the visuomotor task, C. Box-plot of saccadic direction relative to the cursor acceleration during the visuomotor task.

In the framework of specification and implementation of complex digital systems on FPGA, we have developed an approach based on components whose behavior and composition are specified by generalized interpreted T-time Petri nets (HILECOP). One of the inherent difficulties for designer is, on the behavioral part, to account for exceptions. This often leads to a complex modeling and is a source of human errors. Indeed, it is intricate to express all the possible situations (i.e. current state of model). We have defined a way to model exception handling by integrating the well-know concept of macroplace into the formalism. The analysability of the model and the efficiency of the implementation on FPGA (reactivity and surface, ie number of logic blocks) have been preserved. An example of macroplace is given in figure 13; it contains a sub-net (set of places of its refinement) from which exception handling is simply described by a dedicated output transition (transition te on fig. 13), whatever is the current state of the sub-net.

We also solved state evolution conflicts introducing (automatically) priorities between transitions, to avoid reaching inconsistent global state while synchronously executing the model.

The new formalism (including all improvements) has been defined [45], as well as the model transformation based equivalent PNML generation for using existing analysis tools. The VHDL code generation has also been defined [59].

All this work has been applied to an industrial example, that of a neural stimulator developed in collaboration with MXM industrial partner. Results have shown the significant contribution of the theoretical approach to the stimulation device reliability, while preserving both surface and power consumption of the given digital part of the device.

Ongoing work, developed through I. Merzoug PhD thesis, concerns the improvement of the analysis of synchronously implemented Petri nets.

5.3.2. *New FES dedicated digital processor for neurostimulator*

Participants: David Andreu, David Guiraud.

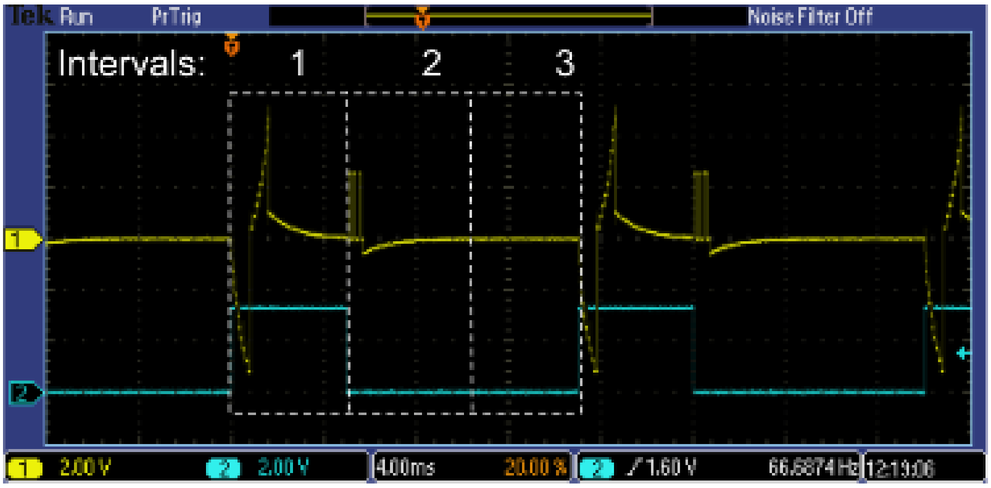
We designed (patent pending) and prototyped a new neural FES dedicated processor and its associated (more compact and efficient) set of instructions, as well as an embedded sequencer for accurate timing in sequencing stimulations to be performed (by the stimulator). The new neural stimulator is based on a dedicated ASIC (Application Specific Integrated Circuit), that is able to drive up to 24 channels of stimulation in absolute synchronization, and with a programmable and controlled current level distribution (patent pending). This ASIC also allows for impedance measurement. The functions of the stimulator are currently implemented in two separate chips: an analog stimulation front-end (ASIC) and a field-programmable gate array (FPGA) embedding the logic control. The FPGA embeds the new FES dedicated processor setting the output stage configuration (poles configuration and current ratio between them) and running potentially complex stimulation profiles (with a 1μs time step and 5 μA current step); example of generated stimulations are shown in Fig.16 . It also embeds the protocol stack allowing for remote programming and online control. Online control relies on advanced and efficient modulation mechanisms, e.g. coefficient based modulation preserving balanced stimulation.

And last but not least, it also embeds a monitoring module ensuring the respect of safety constraints stemming both from target tissue protection and electrode integrity preservation; this reference model based monitoring module ensures (configurable) current and quantity of injected charges limits and thus safe stimulation whatever are electrodes to be used (particularly for thin-film micro-electrodes). Safety limits must be defined by users (partners) according to the target and electrodes to be used.

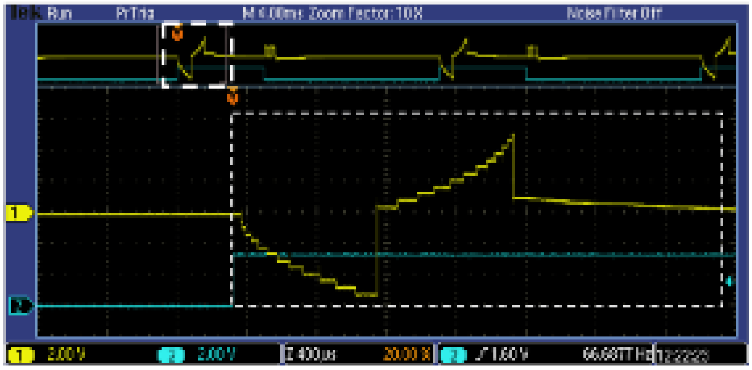
5.3.3. *Fast simulation of hybrid dynamical systems*

Participants: Abir Ben Khaled [IFPEN], Daniel Simon, Mongi Ben Gaid [IFPEN].

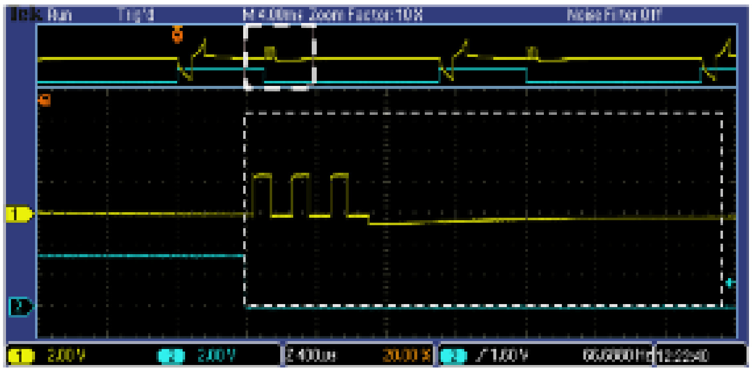
When dealing with the design of complex systems, simulation is an indisputable step between concept design and prototype validation. Realistic simulations allow for the preliminary evaluation, tuning and possibly redesign of proposed solutions ahead of implementation, thus lowering the risks. To be confident in the result, building such simulations needs high fidelity models both for the components and for their interaction. Models



Sequence of stimulation with 3 intervals



zoom on the stimulation profile generated at interval 1



zoom on the stimulation profile generated at interval 2

Figure 16. Sequence of stimulation generated by the new neurostimulator

of dynamical systems (as, for example, muscular fibers) are often given as a set of Ordinary Differential Equations (ODEs). However, the simulation of high-fidelity models is time consuming, and reaching real-time constraints is out of the capabilities of monotithic simulations running on single cores.

The aim of the on-going work is to speed up the numerical integration of hybrid dynamical systems, eventually until reaching a real-time execution, while keeping the integration errors inside controlled bounds. The basic approach consists in splitting the system into sub-models, which are integrated in parallel. I has been shown that an efficient partition must minimize the interactions between sub-models, in particular by confining discontinuities processing inside each component. Automatic partitioning, based on some particular incidence matrices of the original system, has been investigated to introduce a finely-grained co-simulation method enabling numerical integration speed-ups [14]. It is obtained using a partition across the model into loosely coupled sub-systems with sparse communication between modules. The proposed scheme leads to schedule a large number of operations with a wide range of execution times. A suitable off-line scheduling algorithm, based on the input/output dynamics of the models, is proposed to minimize the simulation errors induced by the parallel execution.The method was tested with an automotive engine model, but it is generic and can be applied to other systems of hybrid ODEs/DAEs, as are large sets of muscular fibers.

However, slack synchronization intervals may generate integration errors in the final result. Rather than using costly small integration and communication steps, an enhanced method uses context-based extrapolation is investigated to improve the trade-off between integration speed-ups, needing large communication steps, and simulation precision, needing frequent updates for the models inputs. the method uses extrapolations of the behavior of the models over the synchronization intervals. Test results on a hybrid dynamical engine model, based on FMI for model exchange, show that well chosen context-based extrapolation allows for significant speed-up of the simulation with negligible computing overheads [37].

5.3.4. Control and scheduling co-design for stimulation systems

Participants: Daniel Simon, David Andreu, Samy Lafnoune [Master2 Robotique].

In the FES distributed system developed by the team and marketed by Vivaltis, external electrodes are sticked on the body to either stimulate a muscle or to measure sensitive information (e.g., EMG). Each electrode is connected to a pod, which can be either a stimulation pod or a measurement pod. These Pods are controlled and coordinated by a controler through wireless radio-frequency (RF) links.

The communication frames between the controler and the pods use frames, each frame contains the receiver address and a code which correspond to the deired action. Simple communication stacks, based on the reduced OSI model, are implemented in each node of the wireless network, and the current version only provides static scheduling of the messages.

In fact static scheduling is not at all optimal when the generation of complex motions needs a tight coordination between several sensors and actuators over closed-loop controllers. The high sensitivity of the RF link with respect to varying networking loads and environment conditions calls for an adaptive scheduling of the messages via the regulation of a QoS criterion. The design of such feedback controllers wil rely on previous work [53], [54].

As any modification of the existing devices can be costly and as experiments involving livings cannot be done easily, a real-time simulation system has been design and implemented. The system includes continuous models of the muscular and skeleton systems, models of the wireless network, simplified communicaion stacks and control code running inside real-time threads. The system is open, running C code inside posix threads under Linux, so that the models can be progressively detailed and enriched when necessary [56].

GALEN Project-Team

6. New Results

6.1. Highlights of the Year

- Handbook of Biomedical Imaging: Methodologies and Clinical Research [38] - co-edited from Nikos Paragios, James Duncan and Nicholas Ayache - has been published from Springer Publishing house.
- Nikos Paragios was admitted as a senior fellow at the Insitut Universitaire de France and has been awarded an IBM Faculty award. He has also been one of the plenary invited lecturers at the IARP International Conference in Pattern Recognition (ICPR'2015, Stockholm).

BEST PAPER AWARD :

[26] **Sparsity Techniques in Medical Imaging (STMI)**. M. MISYRLIS, A. KONOVA, M. BLASCHKO, J. HONORIO, N. ALIA-KLEIN, R. GOLDSTEIN, D. SAMARAS.

6.2. Rounding-based Moves for Metric Labeling

Paticipants: M. Pawan Kumar

Metric labeling is an important special case of energy minimizaton in Markov random fields. While the best known polynomial-time algorithm for the problem is the linear programming (LP) relaxation, in practice it is slow to solve it. In [25], we introduced a new family of efficient move-making algorithms for metric labeling. These algorithms mimic the rounding procedues used for converting a fractional LP solution to a feasible integral solution. Our algorithms provide a matching theoretical guarantee to the LP relaxation, while requiring significantly less computational time.

6.3. Optimizing Average Precision

Paticipants: Puneet Kumar Dokania, Aseem Behl, Pritish Mohapatra, C.V. Jawahar, M. Pawan Kumar

Average precision (AP) is one of the most commonly used measures for ranking. However, due to the inefficiency of optimizing it during learning, a common approach is to use surrogate loss functions such as 0-1 loss. In [27], we proposed a new optimization algorithm for AP-SVM that allows training in a similar time to binary SVM. In [23], we extended the AP-SVM framework to score the samples according to high-order information, as opposed to simple first-order information used in prior work. Finally, in [19], we proposed a novel latent AP-SVM formulation that allows learning from weakly supervised datasets. The advantage of learning with high order and missing information is demonstrated on challenging computer vision problems such as action classification and object detection using standard benchmark datasets.

6.4. Discriminative Training of Deformable Contour Models

Paticipants: Haithem Boussaid, Iasonas Kokkinos and Nikos Paragios

Deformable Contour Models (DCMs) are a main workhorse for medical image analysis - but are not commonly studied from a machine learning perspective. In [21], [20] we haved proposed an integrated machine learning and optimization framework to deploy DCMs in medical image analysis.

Our technical contributions are two-fold: firstly, we use an efficient decomposition-coordination algorithm to solve the optimization problems resulting from Loopy DCMs, by means of the Alternating Direction Method of Multipliers (ADMM); this yields substantially faster convergence than plain Dual Decomposition-based methods.

Secondly, we use structured prediction to exploit loss functions that better reflect the performance criteria used in medical image segmentation. By using the mean contour distance (MCD) as a structured loss during training, we obtain clear test-time performance gains.

We demonstrate the merits of exact and efficient inference with rich, structured models in a large X-Ray image segmentation benchmark, where we obtain systematic improvements over the current state-of-the-art.

6.5. Improved Deformable Part Models for Object Detection

Paticipants: Iasonas Kokkinos, Stavros Tsogkas, Eduard Trulls, Pierre-Andre Savalle, George Papandreou.

In [30] and [36] we have worked on improving the classification accuracy of Deformable Part Models (DPMs) for object detection in two distinct manners. Firstly, in [30] we propose a technique to combine bottom-up segmentation, coming in the form of SLIC superpixels, with sliding window DPM detectors. The merit of our approach lies in 'cleaning up' the low- level features by exploiting the spatial support indicated by segmentation. - tion, for both the root and part filters of DPMs. We use these masks to construct enhanced, background- invariant features to train DPMs. We test our approach on the PASCAL VOC 2007, outperforming the standard DPM in 17 out of 20 classes, yielding an average increase of 1.7AP. Additionally, we demonstrate the robustness of this ap- proach, extending it to dense SIFT descriptors for large dis- placement optical flow.

Secondly, in [36] we have explored the potential of convolutional neural networks as feature extractors for detection with DPMs. In particular, we substitute the Histogram-of-Gradient features of DPMs with Convolutional Neural Network (CNN) features, and demonstrate that we thereby obtain a substantial boost in performance (+14.5 mAP) when compared to the baseline HOG-based models. Some more recent extensions to this work are included in [41] where we explore the potential of explicit scale and aspect ratio search in the context of sliding window detection with CNNs.

6.6. Fine-Grained models of objects and texture

Paticipants: Iasonas Kokkinos, Matthew Blaschko, Stavros Tsogkas, Andrea Vedaldi, Mircea Cimpol, Subhransu Maji, Ross Girshick, Juho Kannala, Esa Rahtu, David Weiss, Ben Taskar, Karen Simonyan.

In [31] and [22] we explore methods for the fine-grained understanding of objects and textures, respectively.

In [22] we introduce a texture dataset that is accompanied by descriptions that capture the essence of the textures in terms of attributes. We explore a broad range of classification techniques for these texture attributes and demonstrate that the learned classifiers help improve generic texture recognition methods.

In [31] we introduce a large-scale dataset of airplanes that is accompanied by thorough human annotations at different levels: airplane types, segment lineouts, attributes, and part descriptions are provided for more than 7000 airplane images. We explore the potential of these rich annotations for the task of constructed fine-grained image descriptions using discriminative training techniques on top of standard image representations (Histogram-of-gradient features).

6.7. Large Scale Video Segmentation

Paticipants: Matthew Blaschko

Spatio-temporal cues are powerful sources of information for segmentation in videos. In [24] we present an efficient and simple technique for spatio-temporal segmentation that is based on a low-rank spectral clustering algorithm. The complexity of graph-based spatio-temporal segmentation is dominated by the size of the graph, which is proportional to the number of pixels in a video sequence. In contrast to other works, we avoid oversegmenting the images into super-pixels and instead generalize a simple graph based image segmentation. Our graph construction encodes appearance and motion information with temporal links based on optical flow. For large scale data sets naïve graph construction is computationally and memory intensive, and has only been achieved previously using a high power compute cluster. We make feasible for the first time large scale graph-based spatio-temporal segmentation on a single core by exploiting the sparsity structure of the problem and a low rank factorization that has strong approximation guarantees.

6.8. Higher Order Graph Matching

Paticipants: Chaohui Wang, Dimitris Samaras, Nikos Paragios

In [42] a generic framework for sparse and dense graph/3D surface matching has been introduced. The framework is endowed with a novel mathematical formulation regarding the matching process along with a novel deformation model. It exploits the power of invariance of higher order clique potentials and through a low to high resolution approach determines optimal correspondences between two sets of 3D points while taking advantage of Mobius tranformation to measure local similarity of shapes/graphs/surfaces. Graph matching of objects undergoing non-rigid deformations along with temporal 3D surface tracking demonstrated the potentials of our method. Inference is solved through an efficient dual decomposition scema.

6.9. Inference of Procedural Grammars from Images

Paticipants: Nikos Paragios

Grammar-like representations are powerful modeling and inference tools in computational vision. In [39] a novel approach towards automatic inference of typology specific building grammars has been introduced. The central idea was to consider that such grammars could be derived through a bottom up approach of common sub-tree reasoning of derivation trees determined through parsing using elementary shape (binary split) grammars. Such an approach performs common subtree reduction within the entire training set and identifies meta-rules (corresponding to the same subtrees) which are then clustered together towards producing a compact, typology specific grammar. Promising results both in terms of grammar compactness as well as in terms of inference demonstrated the potentials of the method that could be used beyond the considered scoped.

6.10. Fully connected CRFs for blood vessel segmentation in retinal images

Paticipants: Matthew Blaschko, José Ignacio Orlando

In [28], we present a novel method for blood vessel segmentation in fundus images based on a discriminatively trained, fully connected conditional random field model. Retinal image analysis is greatly aided by blood vessel segmentation as the vessel structure may be considered both a key source of signal, e.g. in the diagnosis of diabetic retinopathy, or a nuisance, e.g. in the analysis of pigment epithelium or choroid related abnormalities. Blood vessel segmentation in fundus images has been considered extensively in the literature, but remains a challenge largely due to the desired structures being thin and elongated, a setting that performs particularly poorly using standard segmentation priors such as a Potts model or total variation.

6.11. Graph-based Segmentation

Paticipants: Sarah Parisot, Deepak Chittajallu, Ioannis Kakadiaris, Nikos Paragios

In [17] we revisited explicit contour-evolution segmentation methods driven from a graph-based shape prior. Prior knowledge through geometric constraints has been encoded to the model within pair-wise interactions between control points. The segmentation process was driven through an objective function seeking to move the control points towards image locations optimizing the expected visual properties of the organ while satisfying the prior geometric constraints being learned at training. In [18] we have proposed a mathematical formalism for automatic tumor segmentation which was taking advantage of conventional segmentation likelihoods and atlas-based segmentation methods. The central idea was to jointly deform and segment an atlas such that the tumor likelihoods are maximized once projected to the targeted image while relaxing the registration constraints in this area. Furthermore we have endowed to this framework explicit estimation of uncertainties allowing the dynamic sampling of the graph structure resulting on significant speed up of the process while producing quantitative means for the interpretation of the final result.

6.12. Multi-atlas Segmentation

Paticipants: Stavros Alchatzidis, Aristeidis Sotiras, Nikos Paragios

In [33] a novel approach that couples pair-wise deformable registration with multi-atlas segmentation using graphical models was proposed. The method exploits both spaces and seeks to determine the optimal solution which will create the best possible visual agreement between atlases and target image along with their label consistency. The approach optimizes the deformation models and the segmentation labels jointly through an interconnected graph allowing either to relax registration constraints when segmentation labels do indicate or the opposite. The joint optimization of both spaces allowing the “implicit” automatic selection of atlases and therefore improves significantly segmentation performance.

6.13. Higher Order Graph Training throuh Dual Decomposition and Max Margin Principles

Paticipants: Nikos Komodakis, Bo Xiang, Nikos Paragios

In [40] a novel framework based on the structure margin principle was introduced for training higher order graphical models. The idea was to reduce the training of a complex high-order MRF in the parallel training of a series of simple slave MRFs through a principled dual decomposition approach. The theoretical properties of the framework have been studied while the method has been experimentally tested using 2d/3d segmentation problems involving higher order geometric priors that are linear-invariant. The proposed formulation benefits from theoretical guarantees as it concerns performance, computational simplicity while being modular and scalable.

MNEMOSYNE Project-Team

6. New Results

6.1. Overview

Though our view is systemic, our daily research activities are concerned with the design, at a given scale of description, of models of neuronal structures, each concerned with a specific learning paradigm. Of course, a major challenge is to keep in mind the systemic view, to put a specific emphasis on the way each neuronal structure communicates with the rest of the system and to highlight how the learning paradigm interplays with other memory systems.

Among the numerous loops involving the brain, the body and the environment, a basic grid of description corresponds to distinguish “Perception Loops”, the goal of which is to extract from the inner and outer world sensory invariants helpful to identify and evaluate the current state and to make predictions from previous learning, and “Action Loops”, the goal of which is to rely on this sensory, emotional and motivational information to decide, plan and trigger actions for the benefit of the body.

Presently, our team is engaged on the following topics: Concerning perception loops, we are firstly considering the role of the hippocampus and of the posterior cortex in learning high level sensory cues that contribute to pavlovian conditioning in the amygdala. Secondly, we are investigating the role of the thalamus in attentional shifts in the cortex. Concerning Action loops, we are preparing a critical analysis of the current views of the interactions between the prefrontal cortex and the basal ganglia. Finally, we also report here more methodological achievements.

6.2. Pavlovian conditioning

Pavlovian conditioning is an outstanding example of a systemic process involving several cerebral structures and several modes of learning. This year, we have made more precise our model of amygdala [7], with a special emphasis on the variety of its inputs, including by neuromodulation [12]. We have also specifically discussed and contrasted the role of cerebral structures involved in this learning, from the point of view of information processing [11]. In addition to the amygdala, the structures of interest are the hippocampus and the posterior and prefrontal cortex and begin to be investigated in ongoing studies.

6.3. The thalamus is more than a relay

Many recent results in neuroscience indicate that the role of the thalamus in the brain is certainly more important than it used to be considered, particularly concerning its relation with the cortex. Our modeling and bibliographic studies were carried out in the Keops project (*cf.* § 7.2) with our chilean neuroscientist colleagues studying non standard ganglion cells in the retina. Particularly, the PhD work by Carlos Carvajal [1] led us to propose a biologically-founded algorithm for the interplay between the modulatory and driving connections between the thalamus and the cortex, with the strong constraint of proposing a system working on a real visual flow.

6.4. On the computational efficiency of Basal Ganglia models

Many valuable models have been proposed to capture the richness of the fundamental relations between the basal ganglia and prominent brain structures including the prefrontal cortex, the hippocampus and the superior colliculus. To choose among them the mechanisms on which to build the design of the motor pole of our brain-inspired system, a fundamental issue is to evaluate the efficiency of these models in more realistic cases than the ones which are generally considered by the authors [24]. For this reason, we have conducted a systematic study of several basal ganglia computational models to check of their scalability in terms of ation representational space [25]. Unfortunately, we found most of them to not be scalable and some of them to not be reproducible at all.

Another way to explore the computational efficiency of neuronal models is to implement them at lower levels of description. This is currently being done with one model developed in our lab at a level corresponding to a neuronal assembly with a mean activity expressed using a single variable. This mesoscopic approach has been refined to a microscopic scale description level, i.e taking into account individual neurons and synapses. Besides the confirmation of many of the results of initial model with a more detailed formalism, this new model has allowed us to highlight the facilitating role of inhibitory interneurons in the decision-making and action selection processes.

6.5. Distributed Self-Organization

The formation of the sensory homunculus in the primary sensory cortex (SI) is believed to be the result of a dynamic neural self-organization process that starts before birth and lasts for several years, allowing the brain to cope with sensory or brain lesions. The exact neural mechanisms driving this self-organization are not yet known and the role of the somatosensory attention remains unclear in this picture. We thus investigated the influence of somatosensory attention onto the two-dimensional structure of area 3b neuronal receptive fields (RFs) using a computational model [2] based on the dynamic neural field theory. This computational model of SI (area 3b) is able to explain experimental data in the monkey and hypothesizes role for the somatosensory attention in the shaping of SI receptive fields.

NEUROMATHCOMP Project-Team

5. New Results

5.1. Highlights of the Year

Olivier Faugeras received the Okawa prize for his pioneering contributions for computer vision and for computational neuroscience. The ceremony will be held in Tokyo in March 2015.

5.2. Neural Networks as dynamical systems

5.2.1. Heteroclinic cycles in Hopfield networks

Participants: Pascal Chossat, Martin Krupa.

Learning or memory formation are associated with the strengthening of the synaptic connections between neurons according to a pattern reflected by the input. According to this theory a retained memory sequence is associated to a dynamic pattern of the associated neural circuit. In this work we consider a class of network neuron models, known as Hopfield networks, with a learning rule which consists of transforming an information string to a coupling pattern. Within this class of models we study dynamic patterns, known as robust heteroclinic cycles, and establish a tight connection between their existence and the structure of the coupling.

This work is available as and has been submitted to a Journal.

5.2.2. Periodic forcing of stabilized E-I networks: Nonlinear resonance curves and dynamics

Participants: Romain Veltz, Terry Sejnowski [Salk Institute].

Inhibition stabilized networks (ISNs) are neural architectures with strong positive feedback among pyramidal neurons balanced by strong negative feedback from in-hibitory interneurons, a circuit element found in the hippocampus and the primary visual cortex. In their working regime, ISNs produce damped oscillations in the γ-range in response to inputs to the inhibitory population. In order to understand the proper-ties of interconnected ISNs, we investigated periodic forcing of ISNs. We show that ISNs can be excited over a range of frequencies and derive properties of the resonance peaks. In particular, we studied the phase-locked solutions, the torus solutions and the resonance peaks. More particular, periodically forced ISNs respond with (possibly multi-stable) phase-locked activity whereas networks with sustained intrinsic oscilla-tions respond more dynamically to periodic inputs with tori. Hence, the dynamics are surprisingly rich and phase effects alone do not adequately describe the network re-sponse. This strengthens the importance of phase-amplitude coupling as opposed to phase-phase coupling in providing multiple frequencies for multiplexing and routing information.

This work has been submitted to a Journal and is available as [38].

5.3. Mean field approaches

5.3.1. A Large Deviation Principle and an Expression of the Rate Function for a Discrete Stationary Gaussian Process

Participants: Olivier Faugeras, James Maclaurin.

We prove a Large Deviation Principle for a stationary Gaussian process over $\mathbb{R}^b$, indexed by $\mathbb{Z}^d$ (for some positive integers d and b), with positive definite spectral density and provide an expression of the corresponding rate function in terms of the mean of the process and its spectral density. This result is useful in applications where such an expression is needed.

This work has been accepted for publication in Entropy [20].

5.3.2. *A representation of the relative entropy with respect to a diffusion process in terms of its infinitesimal-generator*

Participants: Olivier Faugeras, James Maclaurin.

In this paper we derive an integral (with respect to time) representation of the relative entropy (or Kullback-Leibler Divergence) $R(\mu|P)$, where μ and P are measures on $C([0,T];\mathbb{R}^d)$. The underlying measure P is a weak solution to a Martingale Problem with continuous coefficients. Our representation is in the form of an integral with respect to its infinitesimal generator. This representation is of use in statistical inference (particularly involving medical imaging). Since $R(\mu||P)$ governs the exponential rate of convergence of the empirical measure (according to Sanov's Theorem), this representation is also of use in the numerical and analytical investigation of finite-size effects in systems of interacting diffusions.

This work has been accepted for publication in the Journal Entropy [21].

5.3.3. *Asymptotic description of stochastic networks of rate neurons with correlated synaptic weights*

Participants: Olivier Faugeras, James Maclaurin.

We study the asymptotic law of a network of interacting neurons when the number of neurons becomes infinite. Given a completely connected network of neurons in which the synaptic weights are Gaussian correlated random variables, we describe the asymptotic law of the network when the number of neurons goes to infinity. Unlike previous works which made the biologically unplausible assumption that the weights were i.i.d. random variables, we assume that they are correlated. We introduce the process-level empirical measure of the trajectories of the solutions to the equations of the finite network of neurons and the averaged law (with respect to the synaptic weights) of the trajectories of the solutions to the equations of the network of neurons. The result is that the image law through the empirical measure satisfies a large deviation principle with a good rate function. We provide an analytical expression of this rate function. This work has appeared in the Comptes Rendus de l'Academie des Sciences. Serie 1, Mathematique [22].

We have continued the development, started in [22], of the asymptotic description of certain stochastic neural networks. We use the Large Deviation Principle (LDP) and the good rate function H announced there to prove that H has a unique minimum, a stationary measure on the set of trajectories $\mathcal{T}^{\mathbb{Z}}$. We characterize this measure by its two marginals, at time 0, and from time 1 to T. The second marginal is a stationary Gaussian measure. With an eye on applications, we show that its mean and covariance operator can be inductively computed. Finally we use the LDP to establish various convergence results, averaged and quenched. This work has also appeared in the Comptes Rendus de l'Academie des Sciences. Serie 1, Mathematique [23].

5.3.4. *Asymptotic description of stochastic networks of integrate-and-fire neurons*

Participants: François Delarue [UNS, LJAD], James Inglis [EPIs TOSCA and NeuroMathComp], S Rubenthaler [UNS, LJAD], Etienne Tanré [EPI TOSCA].

J. Inglis, together with F. Delarue (Univ. Nice – Sophia Antipolis), E. Tanré (Inria TOSCA) and S. Rubenthaler (Univ. Nice – Sophia Antipolis) completed their study of the mean-field convergence of a highly discontinuous particle system modeling the behavior of a spiking network of neurons, based on the integrate-and-fire model. Due to the highly singular nature of the system, it was convenient to work with a relatively unknown Skorohod topology. The resulting article [46] has been accepted for publication in *Stochastic Processes and Related Fields*.

5.3.5. *Asymptotic description of stochastic networks of spiking neurons with dendrites*

Participants: James Inglis [EPIs TOSCA and NeuroMathComp], Denis Talay [EPI TOSCA].

J. Inglis and D. Talay introduced in [49] a new model for a network of spiking neurons that attempted to address several criticisms of previously considered models. In particular the new model takes into account the role of the dendrites, and moreover includes non-homogeneous synaptic weights to describe the fact that not all neurons have the same effect on the others in the network. They were able to obtain mean-field convergence results, using new probabilistic arguments.

5.3.6. Asymptotic description of stochastic networks of realistic neurons and synapses

Participants: Mireille Bossy [EPI TOSCA], Olivier Faugeras, Denis Talay [EPI TOSCA].

In this note, we clarify the well-posedness of the limit equations to the mean-field N -neuron models proposed in [1] and we prove the associated propagation of chaos property. We also complete the modeling issue in [1] by discussing the well-posedness of the stochastic differential equations which govern the behaviour of the ion channels and the amount of available neurotransmitters.

This work has been submitted for publication to a Journal and is available as [40].

5.3.7. On the Hamiltonian structure of large deviations in stochastic hybrid systems

Participants: Paul Bressloff [Prof. University of Utah, Inria International Chair], Olivier Faugeras.

We develop the connection between large deviation theory and more applied approaches to stochastic hybrid systems by highlighting a common underlying Hamiltonian structure. A stochastic hybrid system involves the coupling between a piecewise deterministic dynamical system in $\mathbb{R}^d$ and a time-homogeneous Markov chain on some discrete space Γ. We assume that the Markov chain on Γ is ergodic, and that the discrete dynamics is much faster than the piecewise deterministic dynamics (separation of time-scales). Using the Perron-Frobenius theorem and the calculus-of-variations, we evaluate the rate function of a large deviation principle in terms of a classical action, whose Hamiltonian is given by the Perron eigenvalue of a $|\Gamma|$-dimensional linear equation. The corresponding linear operator depends on the transition rates of the Markov chain and the nonlinear functions of the piecewise deterministic system. The resulting Hamiltonian is identical to one derived using path-integrals and WKB methods. We illustrate the theory by considering the example of stochastic ion channels. This work has been submitted for publication to a Journal and is available as [41].

5.4. Neural fields theory

5.4.1. Neural fields with noise

Participants: Olivier Faugeras, James Inglis.

We extend the theory of neural fields which has been developed in a deterministic framework by considering the influence spatio-temporal noise. The outstanding problem that we here address is the development of a theory that gives rigorous meaning to stochastic neural field equations, and conditions ensuring that they are well-posed. Previous investigations in the field of computational and mathematical neuroscience have been numerical for the most part. Such questions have been considered for a long time in the theory of stochastic partial differential equations, where at least two different approaches have been developed, each having its advantages and disadvantages. It turns out that both approaches have also been used in computational and mathematical neuroscience, but with much less emphasis on the underlying theory. We present a review of two existing theories and show how they can be used to put the theory of stochastic neural fields on a rigorous footing. We also provide general conditions on the parameters of the stochastic neural field equations under which we guarantee that these equations are well-posed. In so doing we relate each approach to previous work in computational and mathematical neuroscience. We hope this will provide a reference that will pave the way for future studies (both theoretical and applied) of these equations, where basic questions of existence and uniqueness will no longer be a cause for concern. This work has appeared in the Journal of Mathematical Biology [19].

5.4.2. A center manifold result for delayed neural fields equations

Participants: Romain Veltz, Olivier Faugeras.

Lemma C.1 in [83] is wrong. This lemma is used in the proof of the existence of a smooth center manifold, Theorem 4.4. An additional assumption is required to prove this existence. We spell out this assumption, correct the proofs and show that the assumption is satisfied for a large class of delay functions τ. We also weaken the general assumptions on τ.

This work has been accepted for publication in the SIAM Journal on Mathematical Analysis and is available as [37].

5.5. Spike trains statistics

5.5.1. Exact computation of the Maximum Entropy Potential of spiking neural networks models

Participants: Bruno Cessac [correspondent], Rodrigo Cofre.

Understanding how stimuli and synaptic connectivity influence the statistics of spike patterns in neural networks is a central question in computational neuroscience. Maximum Entropy approach has been successfully used to characterize the statistical response of simultaneously recorded spiking neurons responding to stimuli. But, in spite of good performance in terms of prediction, the fitting parameters do not explain the underlying mechanistic causes of the observed correlations. On the other hand, mathematical models of spiking neurons (neuro-mimetic models) provide a probabilistic mapping between stimulus, network architecture and spike patterns in terms of conditional probabilities. In this paper we build an exact analytical mapping between neuro-mimetic and Maximum Entropy models.

This work has been published in [18] and presented in [44], [29], [30], [27].

5.5.2. Parameter Estimation for Spatio-Temporal Maximum Entropy Distributions: Application to Neural Spike Trains

Participants: Bruno Cessac [correspondent], Hassan Nasser.

We propose a numerical method to learn maximum entropy (MaxEnt) distributions with spatio-temporal constraints from experimental spike trains. This is an extension of two papers, [10] and [4], which proposed the estimation of parameters where only spatial constraints were taken into account. The extension we propose allows one to properly handle memory effects in spike statistics, for large-sized neural networks.

This work has been published in [25] and presented in [44], [29], [30], [27].

5.6. Synaptic plasticity

5.6.1. Large Deviations of an Ergodic Synchoronous Neural Network with Learning

Participants: Olivier Faugeras, James Maclaurin.

In this work we determine a Large Deviation Principle (LDP) for a model of neurons interacting on a lattice $\mathbb{Z}^d$. The neurons are subject to correlated external noise, which is modelled as an infinite-dimensional stochastic integral. The probability law governing the noise is strictly stationary, and we are therefore able to find a LDP for the probability laws Π^n governing the ergodic empirical measure μ^n generated by the neurons in a cube of length $(2n+1)$ as n asymptotes to infinity. We use this LDP to determine an LDP for the neural network model. The connection weights between the neurons evolve according to a learning rule / neuronal plasticity, and these results are adaptable to a large variety of specific types of neural network. This LDP is of great use in the mathematical modelling of neural networks, because it allows a quantification of the likelihood of the system deviating from its limit, and also a determination of which direction the system is likely to deviate. The work is also of interest because there are nontrivial correlations between the neurons even in the asymptotic limit, thereby presenting itself as a generalisation of traditional mean-field models.

This work is available [47] and is under review in a Journal.

5.7. Visual Neuroscience

5.7.1. On the effects on cortical spontaneous activity of the symmetries of the network of pinwheels in visual area V1

Participants: Romain Veltz, Pascal Chossat, Olivier Faugeras.

We consider the problem of describing mathematically the spontaneous activity of V1 by combining several important experimental observations including 1) the organization of the visual cortex into a spatially periodic network of hypercolumns structured around pinwheels, 2) the difference between short-range and long-range intra-cortical connections, the first ones being rather isotropic and producing naturally doubly-periodic patterns by Turing mechanisms, the second one being patchy and 3) the fact that the Turing patterns spontaneously produced by the short-range connections and the network of pinwheels have similar periods. By analyzing the Preferred Orientation (PO) map, we are able to classify all possible singular points of the PO maps (the pinwheels) as having symmetries described by a small subset of the wallpaper groups. We then propose a description of the spontaneous activity of V1 using a classical voltage-based neural field model that features isotropic short-range connectivities modulated by non-isotropic long-range connectivities. A key observation is that, with only short-range connections and because the problem has full translational invariance in this case, a spontaneous doubly-periodic pattern generates a 2-torus in a suitable functional space which persists as a flow-invariant manifold under small perturbations, hence when turning on the long-range connections. Through a complete analysis of the symmetries of the resulting neural field equation and motivated by a numerical investigation of the bifurcations of their solutions, we conclude that the branches of solutions which are stable over an extended set of parameters are those corresponding to patterns with an hexagonal (or nearly hexagonal) symmetry. The question of which patterns persist when turning on the long-range connections is answered by 1) analyzing the remaining symmetries on the perturbed torus and 2) combining this information with the Poincaré-Hopf theorem. We have developed a numerical implementation of the theory that has allowed us to produce the patterns of activities predicted by the theory, the planforms. In particular we generalize the contoured and non-contoured planforms predicted by previous authors and predict the existence of mixed contoured/non-contoured planforms. We also found that these planforms are most likely to be time evolving. This work is available as a preprint [36] and has been submitted to a Journal.

5.7.2. Decoding the retina with the first wave of spikes

Participants: Geoffrey Portelli, John Barrett [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK], Timothée Masquelier [Institut de la Vision, UPMC Université Paris 06, Paris, France], Pierre Kornprobst [correspondent].

How a population of retinal ganglion cells (RGCs) encode the visual scene remains an open question. Several coding strategies have been investigated out of which two main views have emerged: considering RGCs as independent encoders or as synergistic encoders, i.e., when the concerted spiking in a RGC population carries more information than the sum of the information contained in the spiking of individual RGCs. Although the RGCs assumed as independent encode the main information, there is currently a growing body of evidence that considering RGCs as synergistic encoders provides complementary and more precise information. Based on salamander retina recordings, it has been suggested [66] that a code based on differential spike latencies between RGC pairs could be a powerful mechanism. Here, we have tested this hypothesis in the mammalian retina. We recorded responses to stationary gratings from 469 RGCs in 5 mouse retinas. Interestingly, we did not find any RGC pairs exhibiting clear latency correlations (presumably due to the presence of spontaneous activity), showing that individual RGC pairs do not provide sufficient information in our conditions. However considering the whole RGC population, we show that the shape of the wave of first spikes (WFS) successfully encodes for spatial cues. To quantify its coding capabilities, we performed a discrimination task and we showed that the WFS was more robust to the spontaneous firing than the absolute latencies are. We also investigated the impact of a post-processing neural layer. The recorded spikes were fed into an artificial lateral geniculate nucleus (LGN) layer. We found that the WFS is not only preserved but even refined through the LGN-like layer, while classical independent coding strategies become impaired. These findings suggest that even at the level of the retina, the WFS provides a reliable strategy to encode spatial cues.

This work is ongoing and was presented as a poster at CNS 2014. See [26] for more details.

5.7.3. Microsaccades enable efficient synchrony-based visual feature detection by retinal ganglion cells

Participants: Timothée Masquelier [Institut de la Vision, UPMC Université Paris 06, Paris, France], Geoffrey Portelli, Pierre Kornprobst [correspondent].

Fixational eye movements are common across vertebrates, yet their functional roles, if any, are debated [70]. To investigate this issue, we exposed the Virtual Retina simulator [84] to natural images, generated realistic drifts and microsaccades [59], and analyzed the output spike trains of the parvocellular retinal ganglion cells (RGC). We first computed cross-correlograms between pairs of RGC that are strongly excited by the image corresponding to the mean eye position. Not surprisingly, in the absence of eye movements, that is when analyzing the tonic (sustained) response to a static image, these cross-correlograms are flat. Adding some slow drift ($\approx$20 min/s, self- avoiding random walk) creates long timescale (>1s) correlations because both cells tend to have high firing rates for central positions. Adding microsaccades ($\approx 0.5°$ in 25ms, that is $\approx 20°$/s) creates short timescale (tens of ms) correlations: cells that are strongly excited at a particular landing location tend to spike synchronously shortly after the landing. What do the patterns of synchronous spikes represent? To investigate this issue, we fed the RGC spike trains to neurons equipped with spike timing-dependent plasticity (STDP) and lateral inhibitory connections [73]. Neurons self-organized, and each one selected a set of afferents that consistently fired synchronously. We then reconstructed the corresponding visual stimuli by convolving the synaptic weight matrices with the RGC receptive fields. In most cases, we could easily recognize what was learned (e.g. a face), and the neuron was selective (e.g. only responded for microsaccades that landed on a face). Without eye movements, or with only the drift, the STDP- based learning failed, because it needs correlations at a timescale roughly matching the STDP time constants [65]. Microsaccades are thus necessary to generate a synchrony-based coding scheme. More specifically, after each microsaccade landing, cells that are strongly excited by the image corresponding to the landing location tend to fire their first spikes synchronously. Patterns of synchronous spikes can be decoded rapidly – as soon as the first spikes are received – by downstream “coincidence detector” neurons, which do not need to know the landing times. Finally, the required connectivity to do so can spontaneously emerge with STDP. As a whole, these results suggest a new role for microsaccades – to enable efficient visual feature learning and detection thanks to synchronization – that differs from other proposals such as time-to-first spike coding with respect to microsaccade landing times.

This work is ongoing and was presented as a poster at CNS 2014.

5.7.4. A new retina-inspired descriptor for image classification

Participants: Cristina Hilario [Pattern Analysis and Computer Vision, PAVIS, Istituto Italiano di Tecnologia, Genova, Italy], Diego Sona [Pattern Analysis and Computer Vision, PAVIS, Istituto Italiano di Tecnologia, Genova, Italy], Kartheek Medathati, Pierre Kornprobst [correspondent].

Recent studies on the visual system reveal that retina is smarter than scientists believed. One low level processing occurring at the retina is feature extraction, becoming an inspiration to build novel image descriptors for image categorization. However only few methods have taken advantage of this idea, such as FREAK descriptor [51], which consists of a circular grid of a concentric distribution of overlapping receptive fields (RFs) in which average image intensities are compared pairwise. In this work we extended such a descriptor but sticking much more to biological data and models of retina. Each RF in our model is described with a linear-nonlinear model (LN) taking into account inhibitory surrounds with parameters based on biological findings. Based on the activity of retinal ganglion cells, we investigated several methods to define a set of descriptors. The performance of each descriptor was tested on computer vision datasets for texture and scene categorization.

This work is ongoing and was presented as a poster at the 1st Workshop of Visual Image Interpretation in Humans and Machine (VIIHM, EPSRC Network for Biological and Computer Vision in the UK).

5.7.5. Shifting stimulus for faster receptive estimation of ensembles of neurons

Participants: Daniela Pamplona, Bruno Cessac, Pierre Kornprobst [correspondent].

The spike triggered averaged (STA) technique has been widely used to estimate the receptive fields (RF) of sensory neurons [58]. Theoretically, it has been shown that when the neurons are stimulated with a white noise stimulus the STA is an unbiased estimator of the neuron RF (up to a multiplicative constant). The error decreases with the number of spikes at a rate proportional to the stimulus variance [75]. Experimentally, for visual neurons, the standard stimuli are checkerboards where block size is heuristically tuned. This raises difficulties when dealing with large neurons assemblies: When the block size is too small, neuron's response might be too weak, and when it is too large, one may miss RFs. Previously online updating the stimulus in the direction of larger stimulus-neural response correlation [62] or mutual information [69], [68] has been proposed. However, these approaches can not be applied for an ensemble of cells recorded simultaneously since each neuron would update the stimulus in a different direction. We propose an improved checkerboard stimulus where blocks are shifted randomly in space at fixed time steps. Theoretically, we show that the STA remains an unbiased estimator of the RF. Additionally, we show two major properties of this new stimulus: (i) For a fixed block sized, RF spatial resolution is improved as a function of the number of possible shifts; (ii) Targeting a given RF spatial resolution, our method converges faster than the standard one. Numerically, we perform an exhaustive analysis of the performance of the approach based on simulated spiked trains from LNP cascades neurons varying RF sizes and positions. Results show global improvements in the RF representation even after short stimulation times. This makes this approach a promising solution to improve RF estimation of large ensemble of neurons.

This work is ongoing and was submitted to COSYNE 2015.

5.7.6. Shifting stimulus for faster receptive estimation of ensembles of neurons

Participants: Kartheek Medathati, Fabio Solari [University of Genoa - DIBRIS, Italy], Manuela Chessa [University of Genoa - DIBRIS, Italy], Guillaume S. Masson [Institut des Neurosciences de la Timone, Team InVibe], Pierre Kornprobst [correspondent].

Motion estimation has been studied extensively in neurosciences in the last two decades. The general consensus that has evolved from the studies in the primate vision is that it is done in a two stage process involving cortical areas V1 and MT in the brain. Spatio temporal filters are leading contenders in terms of models that capture the characteristics exhibited in these areas. Even though there are many models in the biological vision literature covering the optical flow estimation problem based on the spatio-temporal filters little is known in terms of their performance on the modern day computer vision datasets such as Middlebury. In this paper, we start from a mostly classical feedforward V1-MT model introducing a additional decoding step to obtain an optical flow estimation. Two extensions are also discussed using nonlinear filtering of the MT response for a better handling of discontinuities. One essential contribution of this paper is to show how a neural model can be adapted to deal with real sequences and it is here for the first time that such a neural model is benchmarked on the modern computer vision dataset Middlebury. Results are promising and suggest several possible improvements.

This work is ongoing and was presented as a poster at the 1st Workshop of Visual Image Interpretation in Humans and Machine (VIIHM, EPSRC Network for Biological and Computer Vision in the UK). See [35] for more details.

5.7.7. Exploring the richness of center-surround dynamics: A bifurcation study

Participants: Kartheek Medathati, James Rankin [Center for Neural Sciences, NYU, USA], Guillaume S. Masson [Institut des Neurosciences de la Timone, Team InVibe], Pierre Kornprobst [correspondent].

The balance of excitatory and inhibitory interactions between neurons is one of the characteristic aspects of neural computation. In both neural network and neural field models these interactions have been modeled using center-surround connectivity kernels. Depending on the relative strength of excitation and inhibition these networks have been found to exhibit rich and interesting dynamical behavior. Although many models have been reported in the literature using center-surround connectivity kernels and many experimental studies have shown evidence for changes in observed behavior from winner-take-all to gain control, a thorough bifurcation analysis of these networks in terms of sensitivity of the network to peak strength, discriminability of the peaks and speed of convergence has not been done. In our present work we visit this question in order to identify the parameter regimes where this important switch in the behavior of the network occurs and also establish the trade offs that arise with the choice of a particular connectivity kernel.

This work is ongoing and was presented as a poster at the conference "Nonlinear dynamics and stochastic methods: from neuroscience to other biological applications"

5.7.8. From Habitat to Retina:Neural Population Coding using Natural Movies

Participants: Bruno Cessac [correspondent], Ruben Herzog [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Joaquin Araya [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Michael Pizarro [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Cesar Ravello [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile], Maria Jose Escobar [Universidad Técnica Federico Santa María, Valparaiso, Chile], Adrian Palacios [Centro Interdisciplinario de Neurociencia de Valparaíso, Univ de Valparaíso, 2360102 Valparaíso, Chile].

We use a diurnal rodent retina (O. Degus), which has the advantage of present a 3:1 proportion of rods and cones, respectively, to study the RGC population responses to habitat-based natural stimuli. In order to do this, we have developed a mobile robot that is capable to record movies in the natural habitat of this rodent, simulating both his movements and the eye-ground distance, which allows us to stimulate and record an in vitro retina patch using MEA (multi electrode array) with a sequence of images taken from the animal natural habitat. The analysis of spike statistics has been done using the Enas software to characterize spatio-temporal pairwise correlation with Gibbs distributions. potential constitutes a useful tool for comparing pairwise spatio-temporal correlations between different conditions for the same RGC population. We show that correlated spiking patterns represents a major deviation between White Noise and Natural Movies conditions. We also conclude that population coding for this monophasic OFF RGC population is mostly based on spatial correlation when stimulated with Natural Movies.

This ongoing work has been presented in [48].

...

NEUROSYS Team

6. New Results

6.1. Highlights of the Year

Microscopic action affects mesoscopic and macroscopic action in neural systems. In the context of general anaesthesia, it is not understood how single neuron properties, such as ion-channel conductivities or anesthetic action on neuron receptors, translate to population dynamics and consequently to behavior. The work of Laure Buhry and Axel Hutt [4] proposes a modelling approach how to bridge the microscopic and the mesoscopic scale. The most interesting aspect is that this model bridge allows to extend standard neural field theory on the mesoscopic scale instead of introducing a new model.

In addition, we have developed strong collaborations with medical doctors. First, we have established a collaboration with Dr. Denis Schmartz and Dr. Claude Meistelmann at the *CHU Nancy* to plan and perform well-controlled resting state experiments under propofol anaesthesia. Second, we are in close contact to Jean-Luc Schaff at the *CHU Nancy* (together with Laurent Koessler at *CRAN*) in the context of sleep monitoring. Dr. Schaff has provided us polysomnographic data measured during sleep of insomnia patients.

6.2. From the microscopic to the mesoscopic scale

Participants: Laure Buhry, Axel Hutt, Francesco Giovannini, LieJune Shiau

The Highlight of the Year bridges the microscopic scale and the mesoscopic scale. One partial result has already been used in one of our publications [3] to study the link between population dynamics on the mesoscopic scale and the EEG on the macroscopic scale.

In addition, the work of Francesco Giovannini aims at gaining a better understanding of the effects of anaesthesia on the neural correlates of memory, focusing on how anaesthetics disrupt the interaction between the hippocampus and the cerebral cortex. Studies have shown that these two brain structures exhibit a strong synchronisation of their respective neural activity, when performing memory tasks. Neurophysiology experiments have identified various possible candidate generators for rhythmic activity in the area CA1, CA3 and Dentate Gyrus areas of the hippocampus. However the mechanisms by which cortico- hippocampal synchronisation is elicited, and maintained, are yet to be fully understood. As a first step towards this objective, Francesco obtained a working mathematical model of a biologically plausible hippocampal CA1-3 neural cell, based on the Hodgkin-Huxley neuron, capable of exhibiting long-lasting persistent firing activity when subject to a strong transient stimulus. This behaviour is underlay by an intrinsic membrane current activated by the increase of intracellular Calcium ions, following the discharge of an action potential by the neuron. Our hypothesis is that large ensembles of such persistent-firing neurons could sustain the memory-related rhythmic activity displayed by the hippocampus. In this context, Laure Buhry and Axel Hutt work with LieJune Shiau (University of Houston) on a better understanding of the models used by the community of computational neuroscientists. The goal is to show in which extent models are comparable or interchangeable. We focus on the comparison of oscillatory mechanisms of neuronal populations in different spiking models, especially in the Hodgkin-Huxley and the adaptive exponential integrate-and-fire model.

These latter studies link the two description scales by a bottom-up approach.

Conversely, Axel Hutt and collaboration partners from the University of Noth Carolina - Chapel Hill have analysed Local Field Potentials measured in ferrets prefrontal cortex and visual cortex under anesthesia in a top-down analysis [21]. This data allows to extract network interactions in prefrontal cortex and visual cortex and hence revealing underlying mechanisms in general anaesthesia.

6.3. From the mesoscopic to the macroscopic scale

Participants: Laurent Bougrain, Axel Hutt, Pedro Garcia-Rodriguez, Eric Nichols, Guillaume Serrière, Tamara Tosic, Nicole Voges, Mariia Fedotenkova, Meysam Hashemi, Cecilia Lindig-Leon, Kevin Green, Sébastian Rimbert, Thomas Tassone.
To understand the action of anaesthetic drugs on the EEG-signal observed experimentally, Meysam Hashemi has developed and studied several neural mass models [18], [15], [16], [3]. He has identified the thalamo-cortical loop (TCL) as a possible origin of $\delta-$activity. Since loss of consciousness is accompanied by emerging $\delta-$activity, this work relates the TCL to the loss of consciousness.

Increasing the anaesthetic concentration beyond the point of loss of consciousness, EEG-signals exhibit alternating patterns of high and low activity. This activity is called burst suppression. Since these alternations resemble stochastic jumps between low and high activity resting states, Pedro Garcia-Rodriguez and colleagues are working on a stochastic theory based on neural mass models to describe and reproduce these experimental results. Since the minimum mathematical model for such an effect is two-dimensional and does not exhibit potential dynamics, whereas the majority of literature up to date considers one-dimensional stochastic models obeying potential dynamics, Pedro and colleagues had to develop a new stochastic theory. They can show that the two-dimensional dynamics of the neural mass model can be mapped to a one-dimensional stochastic potential model [14], [13]. This reduction allows to apply standard stochastic theory to describe burst suppression as stochastic transistions. This finding indicates the presence of multiple resting states in the brain and supports a heavily discussed hypothesis on the loss of consciousness.

Biological neural networks are subject to random fluctuations, originating from intrinsic random fluctuations of ions or from external stimulus. The latter neural mass models take into account these fluctuations by assuming additive random input fluctuations. For many decades, these additive fluctuations have been assumed to not affect the stability of the system. However, previous own work has revealed that additive fluctuations tune the stability of nonlinear high-dimensional systems. Since random fluctuations play an important role in the description of neural population dynamics and realistic models consider , it is necessary to study in detail how random fluctuations affect the stability of neural mass models and, hence, how our mathematical model analyses have to be modified. To this end, Axel Hutt and colleagues have performed a stochastic center manifold analysis in a delayed stochastic neural mass model [5] and have found conditions for the stability shift. A first application to delayed stochastic neural fields has revealed how additive random fluctuations may affect EEG-signals [19], [6], however additional detailed mathematical studies and the comparison to experimental data are necessary to affirm the importance of the stochastic effect. Essentially, this work emphasizes to take into account nonlinear noise effects in neural mass and neural field models.

Neural mass models do not consider the spatial extension of neural populations and consequently neglect transmission or interaction delay between neurons at different spatial locations. Taking into account the spatial extension and axonal transmission delay, Axel Hutt and colleagues have shown mathematically [7] how travelling activity fronts propagate through neural tissue and how the fronts properties, such as speed, depend on the neural field properties.

The latter neural field model is embedded in a one-dimensional space. Since biological neural populations in the neocortex are organized in two-dimensional layers or sheets, it is necessary to employ neural field models in two spatial dimensions. This causes both theoretically and numerically problems in the presence of axonal transmission delay. Eric Nichols and colleagues has implemented a recent numerical integration algorithm [8] in the visualization software NeuralFieldSimulator, cf. section5.1 . This software is the basis of numerical bifurcation studies of two-dimensional neural field models [12], [20]. First analytical results [10] show good accordance to numerical results obtained by the NeuralFieldSimulator.

The latter neural field models assume homogeneous spatial interactions, i.e., neural interactions whose strength just depends on the distance between the two neurons. This assumption is strong and not biologically realistic in certain brain areas. In addition, this assumption constrains the model description of recurrent sequences of EEG patterns, which have been found experimentally, e.g., during the emergence from general anaesthesia. Consequently to be able to describe such recurrent EEG-pattern sequences, it is necessary to improve the mathematical description of EEG-patterns. A promising new model has been derived by Axel Hutt and collagues based on heterogeneous neural fields [1]. In order to extract the recurrence EEG-patterns from data, we have extended a recent recurrence analysis technique [2]. The next step will consist in the combination of the heterogeneous neural field model and the results from the recurrence analysis.

Recurrence analysis extracts temporally reccurrent time windows in multi-dimensional datasets. Typical EEG-signals obtained durin surgery under anaesthesia include one electrode and hence a single time series only. To extract recurrence structures of such one-dimensional signals, Mariia Fedotenkova computes the multi-dimensional time-frequency representation of the signal and has worked out the best analysis technique for this step [9]. In the next step she will compute the recurrence plot for a large dataset of 110 patients under surgery (data obtained from University of Auckland).

In order to understand immobility during anaesthesia and how to supervise unconscious patients automatically in hospital emergency rooms, Cecilia Lindig-Leon studies motor imagery and its detection by BCI techniques. Limb movement execution or imagination induce sensorimotor rhythms that can be detected in EEG recordings. Her recent work considers signal power changes in two frequency bands to detect the elicited EEG rebound, i.e. the increasing of synchronization, at the end of motor imageries. The analysis is based on the database 2a of the BCI competition IV and shows that rebound can be stronger over the alpha frequency band (8-12Hz) than the beta frequency band (12-20Hz). She can demonstrate that the analysis of the alpha frequency band improves the detection of the end of motor imageries. In this context, Cecilia has compared intrinsic multi-class classifiers (i.e., one-step methods) with ensembles of two-class classifiers on dataset 2a of the BCI competition IV for motor imagery. Subsequently, she has compared the classical Common Spatial Pattern (CSP) approach and the CSP by Joint Approximate Diagonalization in order to identify whether the latter method represents an outperforming alternative.

Sleep is strongly related to anaesthesia and we have started working on the improvement of sleep monitors. The basic idea is to consider not only EEG-signals but multiple different physiological signals (e.g. heart pulses, electrocardiogram, EEG, respiration cycle, body movements) to classify sleep stages. By virtue of the different signal natures of different physiological signals, it is challenging to put together these so-called multi-modal signals in a single analysis method. To this end, Tamara Tosic and colleagues employ recurrence analysis techniques which allow to estimate time windows exhibiting temporal synchronization between physiological signals [11]. They have developed a method that is based on artificial data sets and Local Field Potentials measured under anaesthesia. In the next step, applications to sleep data (obtained from CHU Nancy) will allow to extract sleep stages and will evaluate the method.

PARIETAL Project-Team

6. New Results

6.1. Highlights of the Year

- Congratulations also to Alex and Daniel Strohmeier for their best paper award at the PRNI 2014 conference: “Improved MEG/EEG source localization with reweighted mixed-norms“.
- Elvis Dohmatob got a honorable mention for the student paper award at PRNI 2014 for the work “Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging”

6.2. Which fMRI clustering gives good brain parcellations?

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Elvis Dohmatob.

Analysis and interpretation of neuroimaging data often require one to divide the brain into a number of regions, or parcels, with homogeneous characteristics, be these regions defined in the brain volume or on on the cortical surface. While predefined brain atlases do not adapt to the signal in the individual subjects images, parcellation approaches use brain activity (e.g. found in some functional contrasts of interest) and clustering techniques to define regions with some degree of signal homogeneity. In this work, we address the question of which clustering technique is appropriate and how to optimize the corresponding model. We use two principled criteria: goodness of fit (accuracy), and reproducibility of the parcellation across bootstrap samples. We study these criteria on both simulated and two task-based functional Magnetic Resonance Imaging datasets for the Ward, spectral and K-means clustering algorithms. We show that in general Ward’s clustering performs better than alternative methods with regard to reproducibility and accuracy and that the two criteria diverge regarding the preferred models (reproducibility leading to more conservative solutions), thus deferring the practical decision to a higher level alternative, namely the choice of a trade-off between accuracy and stability.

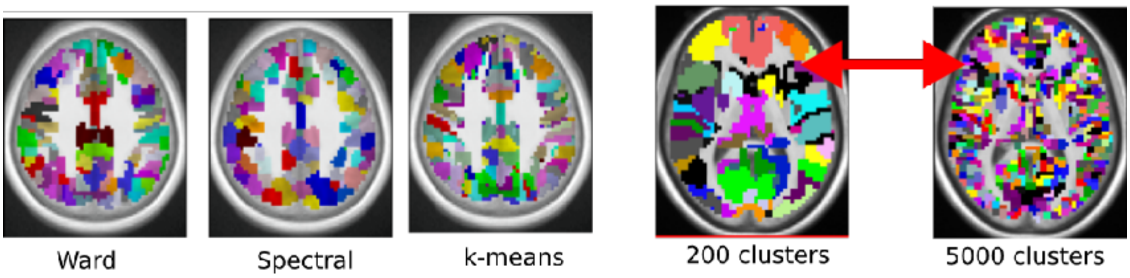


Figure 3. Practitioner have to decide which clustering method to use and how to select the number of clusters. In [21], we provide empirical guidelines and criteria to guide that choice in the context of functional brain imaging.

More details can be found in [21].

6.3. Principal Component Regression predicts functional responses across individuals

Participants: Bertrand Thirion [Correspondant], Gaël Varoquaux, Olivier Grisel.

Inter-subject variability is a major hurdle for neuroimaging group-level inference, as it creates complex image patterns that are not captured by standard analysis models and jeopardizes the sensitivity of statistical procedures. A solution to this problem is to model random subjects effects by using the redundant information conveyed by multiple imaging contrasts. In this paper, we introduce a novel analysis framework, where we estimate the amount of variance that is fit by a random effects subspace learned on other images; we show that a principal component regression estimator outperforms other regression models and that it fits a significant proportion (10% to 25%) of the between-subject variability. This proves for the first time that the accumulation of contrasts in each individual can provide the basis for more sensitive neuroimaging group analyzes.

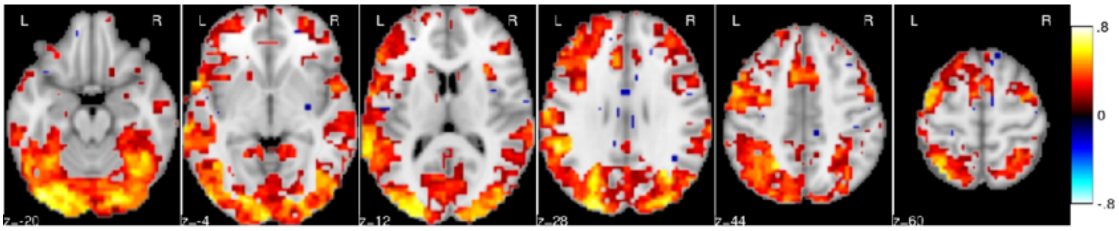


Figure 4. In most brain regions, knowing the amount of activation related to a set of reference contrasts yields an accurate prediction of the activation for a target contrast.

More details can be found in [36].

6.4. Deriving a multi-subject functional-connectivity atlas to inform connectome estimation

Participants: Ronald Phlypo [Correspondant], Bertrand Thirion, Gaël Varoquaux.

The estimation of functional connectivity structure from functional neuroimaging data is an important step toward understanding the mechanisms of various brain diseases and building relevant biomarkers. Yet, such inferences have to deal with the low signal-to-noise ratio and the paucity of the data. With at our disposal a steadily growing volume of publicly available neuroimaging data, it is however possible to improve the estimation procedures involved in connectome mapping. In this work, we propose a novel learning scheme for functional connectivity based on sparse Gaussian graphical models that aims at minimizing the bias induced by the regularization used in the estimation, by carefully separating the estimation of the model support from the coefficients. Moreover, our strategy makes it possible to include new data with a limited computational cost. We illustrate the physiological relevance of the learned prior, that can be identified as a functional connectivity atlas, based on an experiment on 46 subjects of the Human Connectome Dataset.

More details can be found in [35].

6.5. Machine Learning Patterns for Neuroimaging-Genetic Studies in the Cloud

Participants: Virgile Fritsch, Bertrand Thirion, Gaël Varoquaux.

Brain imaging is a natural intermediate phenotype to understand the link between genetic information and behavior or brain pathologies risk factors. Massive efforts have been made in the last few years to acquire high-dimensional neuroimaging and genetic data on large cohorts of subjects. The statistical analysis of such data is carried out with increasingly sophisticated techniques and represents a great computational challenge. Fortunately, increasing computational power in distributed architectures can be harnessed, if new neuroinformatics infrastructures are designed and training to use these new tools is provided. Combining a

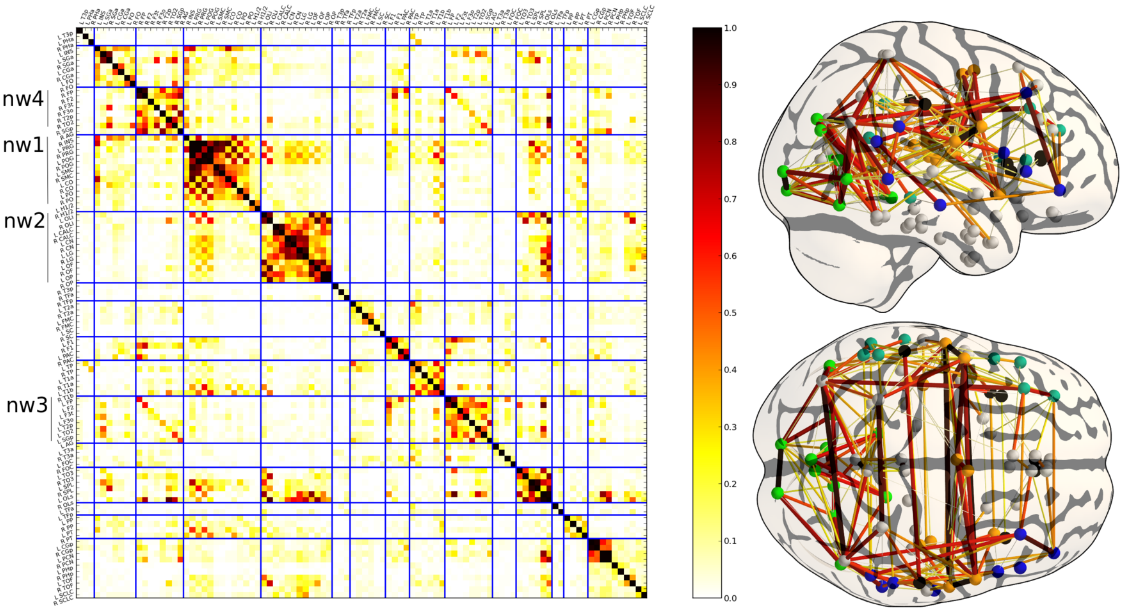


Figure 5. Prior on the functional connectivity: the coefficient of the matrix represent the frequency of an edge at each position. This model can be interpreted as a data-driven atlas of brain functional connections. In the current framework, it can easily be updated to take into account more data.

MapReduce framework (TomusBLOB) with machine learning algorithms (Scikit-learn library), we design a scalable analysis tool that can deal with non-parametric statistics on high-dimensional data. End-users describe the statistical procedure to perform and can then test the model on their own computers before running the very same code in the cloud at a larger scale. We illustrate the potential of our approach on real data with an experiment showing how the functional signal in subcortical brain regions can be significantly fit with genome-wide genotypes. This experiment demonstrates the scalability and the reliability of our framework in the cloud with a two weeks deployment on hundreds of virtual machines.

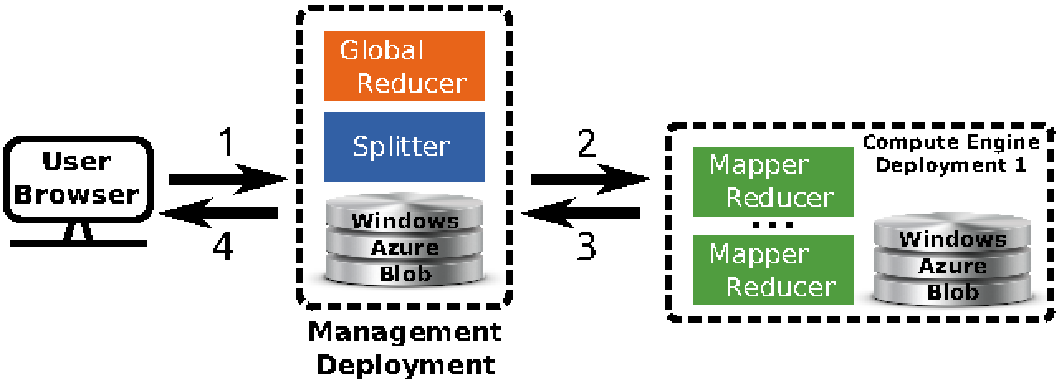


Figure 6. Overview of the multi site deployment of a hierarchical Tomus-MapReduce compute engine. 1) The end-user uploads the data and configures the statistical inference procedure on a webpage. 2) The Splitter partitions the data and manages the workload. The compute engines retrieves job information trough the Windows Azure Queues. 3) Compute engines perform the map and reduce jobs. The management deployment is informed of the progression via the Windows Azure Queues system and thus can manage the execution of the global reducer. 4) The user downloads the results of the computation on the webpage of the experiment.

More details can be found in [17].

6.6. Data-driven HRF estimation for encoding and decoding models

Participants: Fabian Pedregosa Izquierdo [correspondant], Michael Eickenberg, Alexandre Gramfort, Philippe Ciuciu, Bertrand Thirion, Gaël Varoquaux.

Despite the common usage of a canonical, data-independent, hemodynamic response function (HRF), it is known that the shape of the HRF varies across brain regions and subjects. This suggests that a data-driven estimation of this function could lead to more statistical power when modeling BOLD fMRI data. However, unconstrained estimation of the HRF can yield highly unstable results when the number of free parameters is large. We develop a method for the joint estimation of activation and HRF using a rank constraint causing the estimated HRF to be equal across events/conditions, yet permitting it to be different across voxels. Model estimation leads to an optimization problem that we propose to solve with an efficient quasi-Newton method exploiting fast gradient computations. This model, called GLM with Rank-1 constraint (R1-GLM), can be extended to the setting of GLM with separate designs which has been shown to improve decoding accuracy in brain activity decoding experiments. We compare 10 different HRF modeling methods in terms of encoding and decoding score in two different datasets. Our results show that the R1-GLM model significantly outperforms competing methods in both encoding and decoding settings, positioning it as an attractive method both from the points of view of accuracy and computational efficiency.

More details can be found in [19].

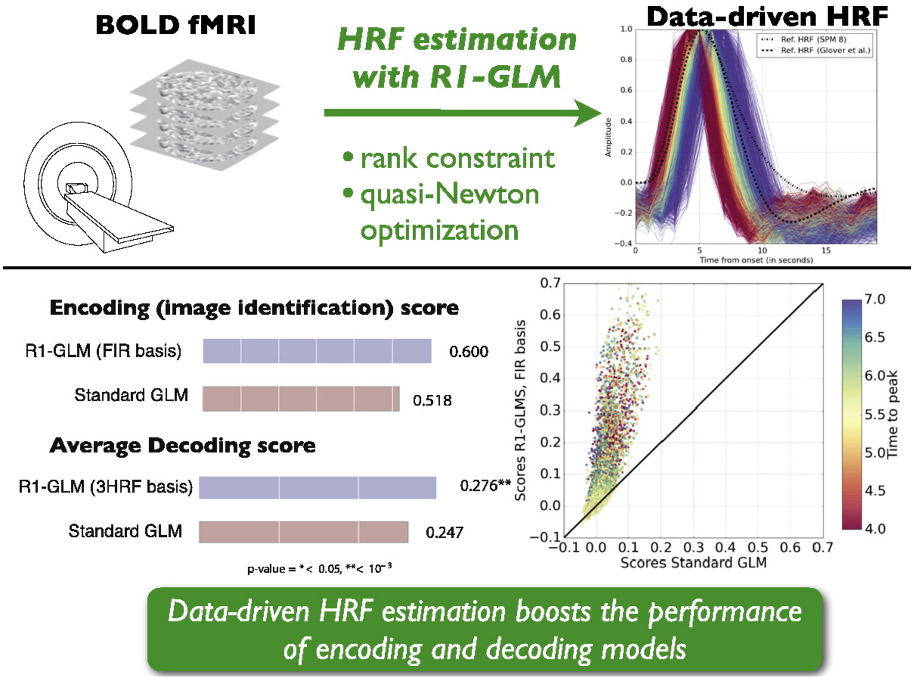


Figure 7. Illustration of the hemodynamic response function estimation framework introduced in [19].

6.7. Benchmarking solvers for TV-l1 least-squares and logistic regression in brain imaging

Participants: Elvis Dohmatob [correspondant], Michael Eickenberg, Gaël Varoquaux, Bertrand Thirion.

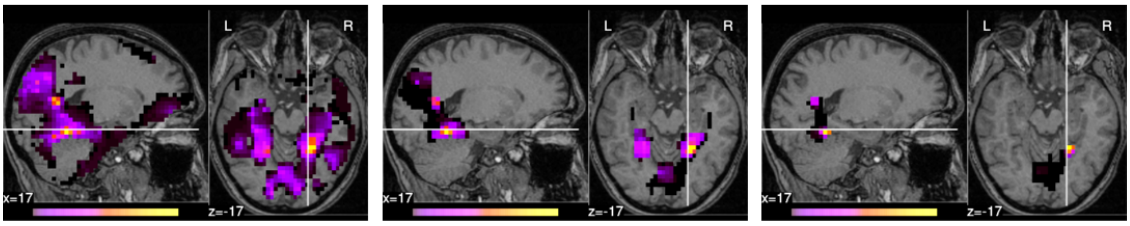
Learning predictive models from brain imaging data, as in decoding cognitive states from fMRI (functional Magnetic Resonance Imaging), is typically an ill-posed problem as it entails estimating many more parameters than available sample points. This estimation problem thus requires regularization. Total variation regularization, combined with sparse models, has been shown to yield good predictive performance, as well as stable and interpretable maps. However, the corresponding optimization problem is very challenging: it is non-smooth, non-separable and heavily ill-conditioned. For the penalty to fully exercise its structuring effect on the maps, this optimization problem must be solved to a good tolerance, resulting in a computational challenge. In this work, we explore a wide variety of solvers and exhibit their convergence properties on fMRI data. We introduce a variant of smooth solvers and show that it is a promising approach in these settings. Our findings show that care must be taken in solving TV-l1 estimation in brain imaging and highlight the successful strategies.

More details can be found in [30]

6.8. Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks

Participant: Philippe Ciuciu [correspondant].

Studies employing functional connectivity-type analyses have established that spontaneous fluctuations in functional magnetic resonance imaging (fMRI) signals are organized within large-scale brain networks. Meanwhile, fMRI signals have been shown to exhibit 1/f-type power spectra – a hallmark of scale-free



$\Delta E = .1$ $\Delta E = .001$ $\Delta E = 10^{-5}$

Figure 8. TV-l1 maps for a face-house discrimination task taken from a visual recognition dataset, with regularization parameters chosen by cross-validation, for different stopping criteria. Note that the stopping criterion is defined as a threshold on the energy decrease per one iteration of the algorithm. This figure shows the importance of convergence of the multivariate estimator, and motivates the need for a fast solver.

dynamics. We studied the interplay between functional connectivity and scale-free dynamics in fMRI signals, utilizing the fractal connectivity framework – a multivariate extension of the univariate fractional Gaussian noise model, which relies on a wavelet formulation for robust parameter estimation. We applied this framework to fMRI data acquired from healthy young adults at rest and performing a visual detection task. First, we found that scale-invariance existed beyond univariate dynamics, being present also in bivariate cross-temporal dynamics. Second, we observed that frequencies within the scale-free range do not contribute evenly to inter-regional connectivity, with a systematically stronger contribution of the lowest frequencies, both at rest and during task. Third, in addition to a decrease of the Hurst exponent and inter-regional correlations, task performance modified cross-temporal dynamics, inducing a larger contribution of the highest frequencies within the scale-free range to global correlation.

More details can be found in [16].

6.9. Supramodal processing optimizes visual perceptual learning and plasticity

Participants: Philippe Ciuciu [correspondant], Alexandre Gramfort.

Multisensory interactions are ubiquitous in cortex and it has been suggested that sensory cortices may be supramodal i.e. capable of functional selectivity irrespective of the sensory modality of inputs. Here, we asked whether learning to discriminate visual coherence could benefit from supramodal processing. To this end, three groups of participants were briefly trained to discriminate which of a red or green intermixed population of random-dot-kinematograms (RDKs) was most coherent in a visual display while being recorded with magnetoencephalography (MEG). During training, participants heard no sound (V), congruent acoustic textures (AV) or auditory noise (AVn); importantly, congruent acoustic textures shared the temporal statistics – i.e. coherence – of visual RDKs. After training, the AV group significantly outperformed participants trained in V and AVn although they were not aware of their progress. In pre- and post-training blocks, all participants were tested without sound and with the same set of RDKs. When contrasting MEG data collected in these experimental blocks, selective differences were observed in the dynamic pattern and the cortical loci responsive to visual RDKs. First and common to all three groups, vlPFC showed selectivity to the learned coherence levels whereas selectivity in visual motion area hMT+ was only seen for the AV group. Second and solely for the AV group, activity in multisensory cortices (mSTS, pSTS) correlated with post-training performances; additionally, the latencies of these effects suggested feedback from vlPFC to hMT+ possibly mediated by temporal cortices in AV and AVn groups. Altogether, we interpret our results in the context of the Reverse Hierarchy Theory of learning in which supramodal processing optimizes visual perceptual learning by capitalizing on sensory-invariant representations - here, global coherence levels across sensory modalities.

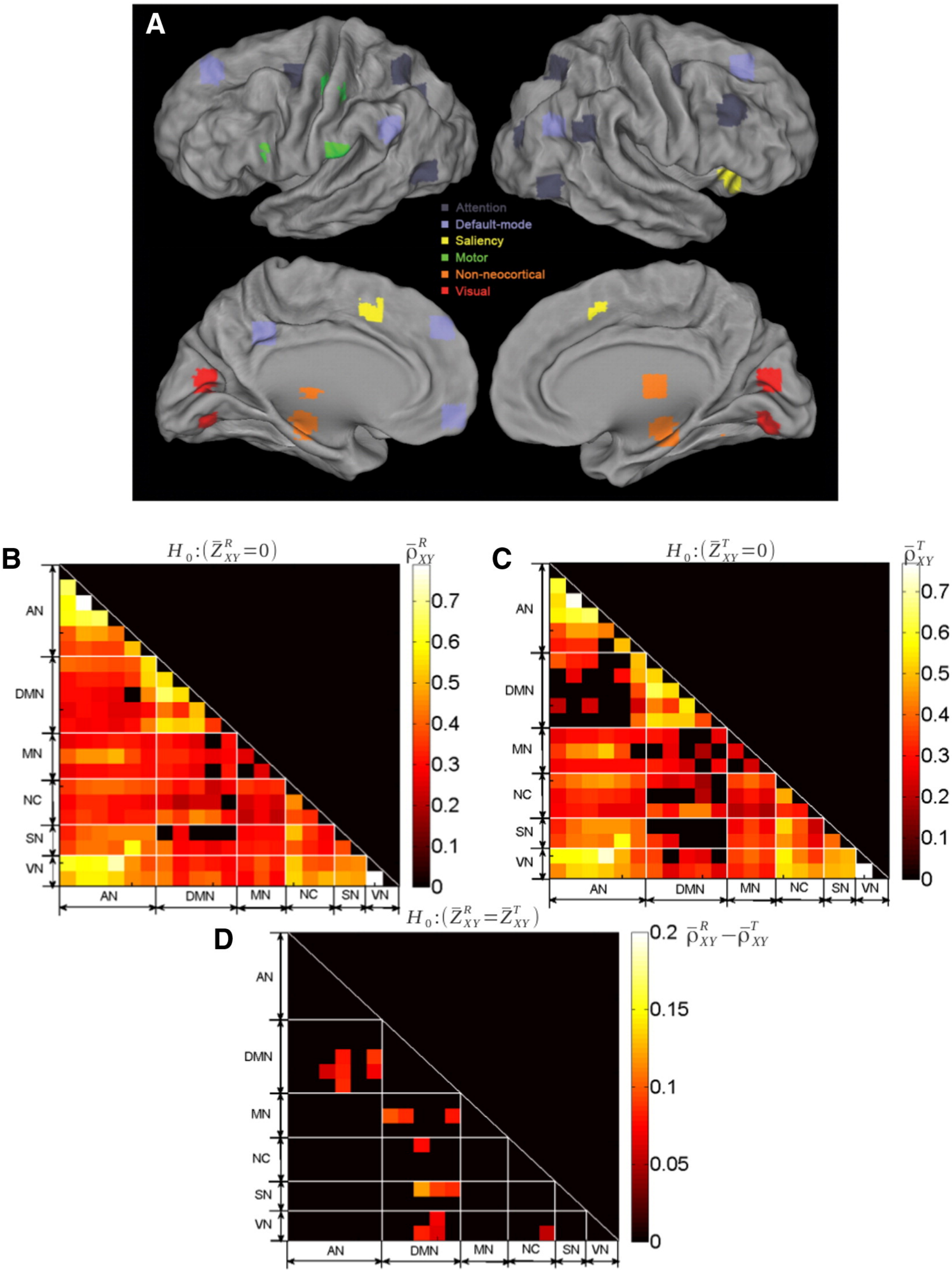


Figure 9. Networks definition and correlation structure. Top (A): ROIs mapped onto the cortical surface, with each color denoting a different network. Middle (B): Group-averaged inter-regional correlation matrix at rest ($p < 0.05$, Bonferroni corrected). Regions are grouped by network to ease visualization. Middle (C): Group-averaged inter-regional correlation matrix during the visual detection task ($p < 0.05$, Bonferroni corrected). Bottom (D): Difference of the correlation coefficients between rest and task (thresholded at $p < 0.01$, uncorrected, two-sample t-test for rest vs. task). The ROIs are grouped by networks; these networks correspond to the diagonal triangles surrounded by white dashed lines.

More details can be found in [25].

6.10. Variable density sampling with continuous trajectories. Application to MRI.

Participants: Nicolas Chauffert, Philippe Ciuciu [correspondant].

Reducing acquisition time is a crucial challenge for many imaging techniques. Compressed Sensing (CS) theory offers an appealing framework to address this issue since it provides theoretical guarantees on the reconstruction of sparse signals by projection on a low dimensional linear subspace. In this paper, we focus on a setting where the imaging device allows to sense a fixed set of measurements. We first discuss the choice of an optimal sampling subspace (smallest subset) allowing perfect reconstruction of sparse signals. Its standard design relies on the random drawing of independent measurements. We discuss how to select the drawing distribution and show that a mixed strategy involving partial deterministic sampling and independent drawings can help breaking the so-called "coherence barrier". Unfortunately, independent random sampling is irrelevant for many acquisition devices owing to acquisition constraints. To overcome this limitation, the notion of Variable Density Samplers (VDS) is introduced and defined as a stochastic process with a prescribed limit empirical measure. It encompasses samplers based on independent measurements or continuous curves. The latter are crucial to extend CS results to actual applications. Our main contribution lies in two original continuous VDS. The first one relies on random walks over the acquisition space whereas the second one is heuristically driven and rests on the approximate solution of a Traveling Salesman Problem. Theoretical analysis and retrospective CS simulations in magnetic resonance imaging highlight that the TSP-based solution provides improved reconstructed images in terms of signal-to-noise ratio compared to standard sampling schemes (spiral, radial, 3D iid...).

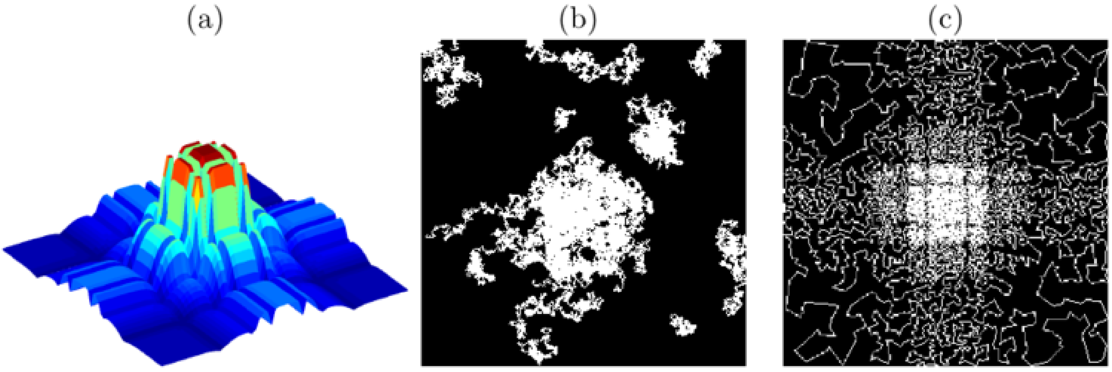


Figure 10. (a): Target distribution π to be approximated. Continuous random trajectories reaching distribution π based on Markov chains (b) and on a TSP solution (c). The latter is much more accurate.

More details can be found in [15].

POPIX Team

6. New Results

6.1. Highlights of the Year

Marc Lavielle published the book, *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools* (Chapman & Hall/CRC), which presents a rigorous framework for describing, implementing, and using mixed effects models. With these models, readers can perform parameter estimation and modeling across a whole population of individuals at the same time.

6.2. New result 1

We have proposed a nonlinear mixed-effects framework to jointly model longitudinal and repeated time-to-event data. A parametric nonlinear mixed-effects model is used for the longitudinal observations and a parametric mixed-effects hazard model for repeated event times. We have shown the importance for parameter estimation of properly calculating the conditional density of the observations (given the individual parameters) in the presence of interval and/or right censoring. Parameters are estimated by maximizing the exact joint likelihood with the Stochastic Approximation Expectation-Maximization algorithm. This workflow for joint models is now implemented in the Monolix software, and illustrated on several simulated and real data examples.

6.3. New result 2

We have succesfully extended the methodologies previously developed for ordinary differential equations (ODE) to delay differential equations (DDE). A C++ solver for DDE, and based on an explicit Runge-Kutta scheme, has been developed. This solver can now be used with Monolix, a platform for population modeling of longitudinal data, MlxPlore, a tool for the exploration of complex models and Simulx a R and Matlab function for the simulation of longitudinal data. We use.

SHACRA Project-Team

5. New Results

5.1. Highlights of the Year

5.1.1. Intra-operative guidance

Each year in Europe 50,000 new liver cancer cases are diagnosed for which hepatic surgery combined to chemotherapy is the most common treatment. In particular the number of laparoscopic liver surgeries has increased significantly over the past years. Minimally invasive procedures are challenging for the surgeons due to the limited field of view.

Providing new solutions to assist surgeons during the procedure is of primary interest. This year, the team developed an innovative system for augmented reality in the scope of minimally invasive hepatic surgery. The first issue is to align preoperative data with the intra-operative images. We first proposed a semi-automatic approach [28] for solving the ill-posed problem of initial alignment for augmented reality systems during liver surgery. Our registration method relies on anatomical landmarks extracted from both the laparoscopic images and a three-dimensional model, using an image-based soft-tissue reconstruction technique and an atlas-based approach, respectively.

Second, we introduced a method for tracking the internal structures of the liver during robot-assisted procedures [25]. Vascular network, tumors and cut planes, computed from pre-operative data, can be overlaid onto the laparoscopic view for image-guidance, even in the case of large motion or deformation of the organ. This is made possible by relying on a fast yet accurate 3D biomechanical model of the liver combined with a robust visual tracking approach designed to properly constrain the model. Our augmented reality proved to be accurate and extremely promising on in-vivo sequences of a human liver during robotic surgery.

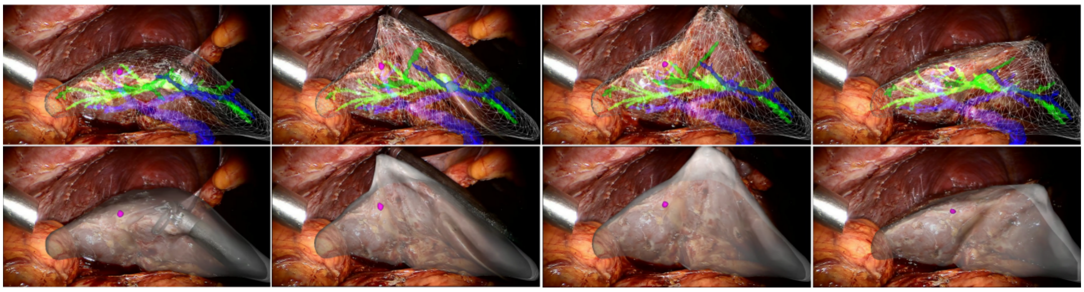


Figure 4. Augmented reality on the liver with 3D visualization of the blood vessels

5.1.2. Ph.D. defenses

The year 2014 was also special since many PhDs have been defended. Four PhD defenses took place with:

- Ahmed Yureidini's defense about *Robust blood vessel surface reconstruction for interactive simulations from patient data* [15] in May 2014,
- Guillaume Kazmitcheff's defense about *Minimal invasive robotics dedicated to otological surgery* [13] in June 2014,
- Hugo Talbot's defense about *Interactive patient-specific simulation of cardiac electrophysiology* [14] in July 2014,
- Alexandre Bilger's defense about *Patient-specific biomechanical simulation for deep brain stimulation* [12] in December 2014.

5.1.3. Organization of ISBMS 2014

The team co-organized the 6th International Symposium on Biomedical Simulation (ISBMS) 2014, which was held in Strasbourg (France) on October 16 – 17, 2014. The ISBMS conference is a well-established scientific meeting that provides an international forum for researchers interested in using biomedical simulation technology for the improvement of patient care and patient safety. The SiMMS group from Imperial College London and IHU-Strasbourg were the two other co-organizers. The event was hosted at IRCAD, a center of excellence in surgical training. The ISBMS chairs were:

- Stéphane Cotin (Inria),
- Fernando Bello (Imperial College London),
- Jérémie Dequidt (Univ. Lille),
- Igor Peterlik (IHU Strasbourg & Masaryk Univ.).

The whole team was involved in the organization of the event. About 65 participants joined the conference. Regarding their feedback, the conference was a real success. For more information about ISBMS, refer to the official website http://www.isbms.org.

Finally, a day dedicated to our software SOFA ("SOFA Day") was organized the day after the ISBMS conference. This was the opportunity to introduce SOFA to the ISBMS community and to share with the SOFA users.

(a) Setup of our demo (b) With Genevieve Fioraso

Figure 5. Presentation of our work at the French National Assembly. Genevieve Fioraso is the French national research secretary

5.1.4. Demonstration at the French National Assembly

On Tuesday 21st January 2014, the team SHACRA presented its work during the "Internet et société numérique" working group. This was a joint event between Inria and the French National Assembly (Assemblée Nationale). On this special occasion, we made a demonstration of our simulations and the CEO from Inria Michel Cosnard also presented more globally the role of Inria in healthcare but also education, cloud computing, big data.

5.2. New Results

5.2.1. Real-Time Biophysical Models

5.2.1.1. Deep brain stimulation

Participant: Alexandre Bilger.

During this year, we developed an intra-operative registration method. It is used during a DBS surgery and can help the surgeon to locate anatomical structures for a safer and a more efficient treatment [21]. The method is based on the biomechanical model of brain shift we developed during the last years. Because some parameters of the model are unknown, we propose to estimate them with an optimization process. The cost function evaluates the distance between the model and the segmentation of pneumocephalus, the only indicator of brain shift visible on an intra-operative CT scan.

Figure 6. Biomechanical model of the brain for DBS planning

5.2.1.2. Stapedectomy

Participant: Guillaume Kazmitcheff.

Stapedectomy is a challenging procedure of the middle ear microsurgery, since the surgeon is in direct contact with sensitive structures such as the ossicular chain. This procedures is taught and performed in the last phase of the surgical apprenticeship. To improve surgical teaching, we propose to use a virtual surgical simulator [26] based on a finite element model of the middle ear. The static and dynamic behavior of the developed finite element model was successfully compared to published data on human temporal bones specimens. A semi-automatic algorithm was developed to perform a quick and accurate registration of our validated mechanical atlas to match the patient dataset. This method avoids a time-consuming work of manual segmentation, parameterization, and evaluation. A registration is obtained in less than 260 seconds with an accuracy close to a manual process and within the imagery resolution. The computation algorithms, allowing carving, deformation of soft and hard tissus, and collision response, are compatible with a real-time interactive simulation of a middle ear procedure. As a future work, we propose to investigate new robotized procedures of the middle ear surgery in order to develop new applications for the RobOtol device and to provide a training tool for the surgeons.

5.2.1.3. Cardiac electrophysiology

Participant: Hugo Talbot.

Cardiac arrhythmia is a very frequent pathology that comes from an abnormal electrical activity in the myocardium. The skills required for such interventions are still very challenging to learn, and typically acquired over several years. We first developed a training simulator for interventional electrocardiology and thermo-ablation of these arrhythmias [14], [32]. Based on physical models, this training system reproduces the different steps of the procedure, including endovascular navigation, electrophysiological mapping, pacing and cardiac ablation. Based on a scenario of cardiac arrhythmia, cardiologists assessed the interactivity and the realism of our simulation.

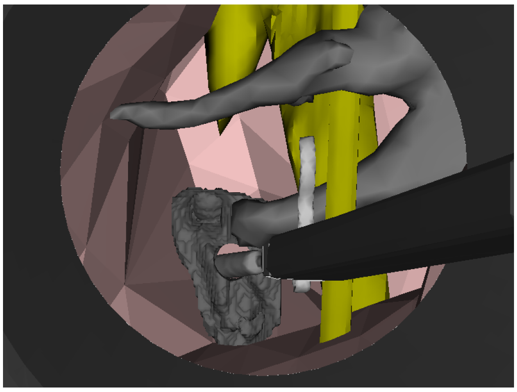


Figure 7. Simulation of the stapedotomy procedure

Figure 8. Training simulator for electrocardiology procedures

Beyond electrophysiology training, our work around the cardiac electrophysiology also target the personalization of our mathematical models. Using the dense electrograms recorded intra-operatively, we presented an accurate and innovative approach to personalize our model, i.e. estimate patient-specific parameters. The modeling in silico of a patient electrophysiology is needed to better understand the mechanism of cardiac arrhythmia. This work has been submitted in a conference.

5.2.1.4. *Cryoablation*

Participant: Hugo Talbot.

A new project started this year around cryotherapy. This technique consists in inserting needles that freezing the surrounding tissues, thus immediately leading to cellular death of the tissues. Cryoablation procedure is used in many medical fields for tumor ablation, and even starts being used in cardiology. In this scope, we build a simulator able to place the cryoprobes and run a simulation representing the evolution of iceballs in living tissues [31]. This work was presented at MMVR'14.

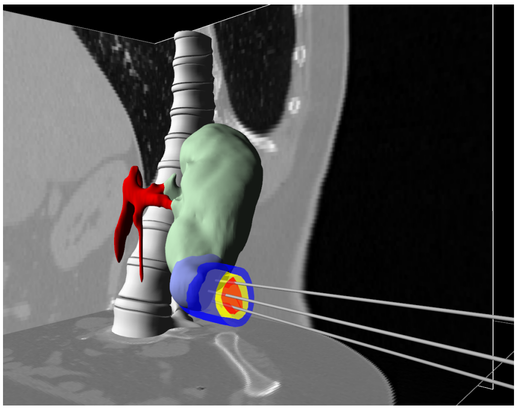


Figure 9. Simulation of the stapedotomy procedure

5.2.1.5. *Connective tissues*

Participant: Julien Bosman.

Another topic of simulation is the modeling of connective tissues [18]. First, a comparative study on the influence of the ligaments in liver surgery has been conducted. This study underlines that the model chosen for the ligament's has a strong influence on the outcome of the simulation. More specifically, it shows the the model is at least as much important as the material parameters of the parenchyma. It also shows that the influence of the model depends on the type of effort that is prescribed on the liver. The second axis concerns the validation of a frame (6-DOF nodes) based mechanical model developed for ligaments simulation. Current results show that this model requires less degrees of freedom while providing the same accuracy as a traditional FEM model. At last, a method dedicated to the simulation and the control of continuum robots has been developed. The goal of this method is to replace the mesh of robot by computing its compliance and applying it on a reduced model made of frames. It allows to strongly decrease the number of degrees of freedom needed for the robot simulation while keeping the needed accuracy.

5.2.1.6. *Simulation of lipofilling reconstructive surgery*

Participant: Vincent Majorczyk.

We have developed a method to simulate the outcome of reconstructive facial surgery based on fat-filling. Facial anatomy is complex: the fat is constrained between layers of tissues which behave as walls along the face; in addition, connective tissues that are present between these different layers also influence the fat-filling procedure. To simulate the end result, we have proposed a method which couples a 2.5D Eulerian fluid model for the fat and a finite element model for the soft tissues. The two models are coupled using the computation of the mechanical compliance matrix. We had two contributions: a solver for fluids which couples properties of solid tissues and fluid pressure, and an application of this solver to fat-filling surgery procedure simulation.

5.2.1.7. Inverse FEM simulation

Participant: Eulalie Coevoet.

We introduced a new methodology for semi-automatic deformable registration of anatomical structures [23], using interactive inverse simulations. We applied the approach for the registration of the parotid glands during the radiotherapy of the head and neck cancer. Radiotherapy treatment induces weight loss that modifies the shape and the positions of these structures and they eventually intersects the target volume. We proposed a method to adapt the planning to limit the radiation of these glands.

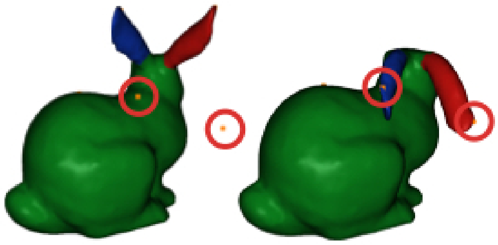


Figure 10. Numerical validation. Left: Target points (highlighted in red) after setting 3 different Young's moduli (one color by Young's modulus). Right: The resulting deformation once the Young modulus have been estimated.

5.2.2. Numerical Methods for Complex Interactions

5.2.2.1. Cliping in neurosurgery

Participant: Eulalie Coevoet.

We developed a simulator for neurosurgery. The surgery consist in "clipping" a cerebral aneurysm. Aneurysm is an abnormal local dilatation in the wall of a blood vessel, usually an artery. There are several treatment options for people with the diagnosis of cerebral aneurysm. Medical therapy, surgical therapy (clipping) and endovascular therapy (coiling). The surgical therapy, because of his invasive and technical nature, is the less prescribed. This leads to less and less surgeon trained to practice the procedure. And yet some patients require the surgical way. So the idea was to develop a simulator to train student and also help on the planification.

5.2.2.2. Virtual cutting

Participants: Huu Phuoc Bui, Christoph Paulus.

The simulation of cutting is a central interest in the team. Several approaches have been investigated this year to model surgical cuts, tearing and other separations of materials induced by surgical tools:

- using the standard finite element method (FEM) combined with a re-meshing approach, that replaces locally the current structure of the mesh in order to allow for a separation,
- using the extended FEM (X-FEM) that uses shape functions that can model discontinuities inside elements (see Fig. 12),
- and using the Lattice element method (LEM).

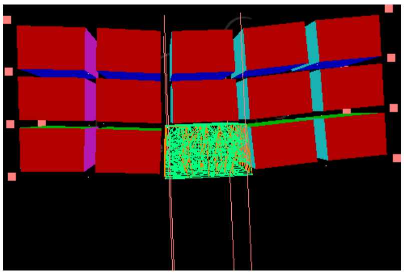
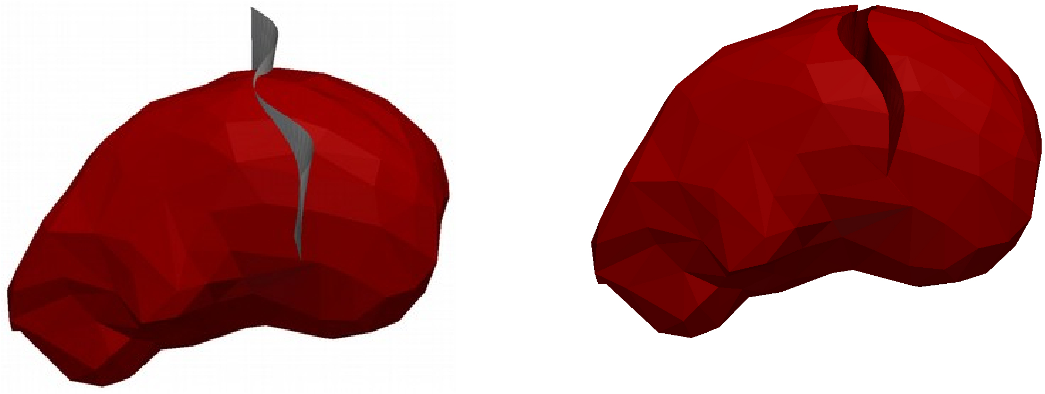


Figure 11. Cutting simulation using LEM

A re-meshing approach to model cuts has been submitted to several conferences, we are waiting for the response. An implementation of the extended finite element method was published in a preprint "Simulation of Complex Cuts in Soft Tissue with the Extended Finite Element Method (X-FEM)". The figures below show a simulation of a sinusoidal cut on a liver executed with the implementation of the X-FEM.

For the LEM approach (see Fig. 11), a multimapping between finite elements and lattice model have been developed and implemented into SOFA framework. This allows us to perform a multiscale simulation in real-time. A dynamic changing of topology between finite elements and lattices should be developed in the next step in order to perform the cutting dynamically.



(a) Plane of cut

(b) Result of the cutting

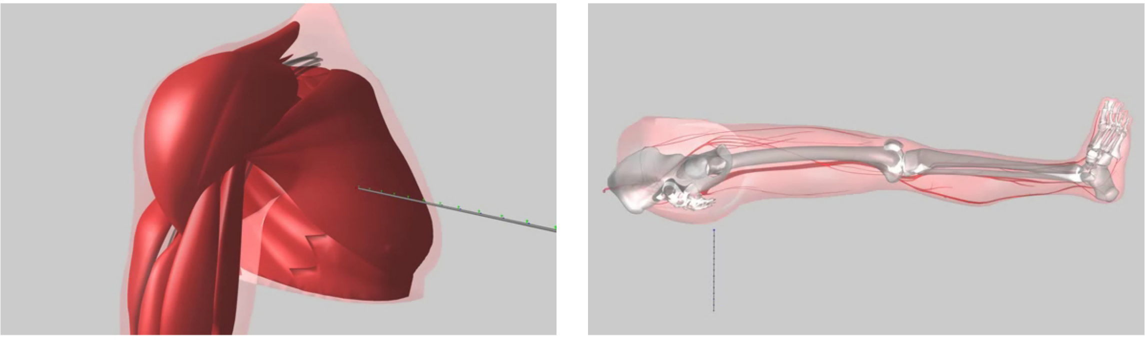
Figure 12. Cutting simulation based on X-FEM

5.2.2.3. Regional anaesthesia

Participants: Rémi Bessard Duparc, Frédérick Roy.

The RASimAs project (Regional Anaesthesia Simulator and Assistant) is a European research project funded by the European Union's 7th Framework Program. It aims at providing a virtual reality simulator and assistant to doctors performing regional anaesthesia by developing the patient-specific Virtual Physiological Human models. This year, the code for needle insertion has been re-designed and simplified into SOFA and the muscle contraction has been implemented. Finally, the components of the simulation have been optimized to reach the desired real-time performances (i.e more than 25-30 frames per second).

Our preliminary results are awaiting the validation of the Working Packages in January 2015. The needle refactoring will be shared with an other project in Strasbourg (robot) and may be shared with an other team at Inria Rennes with the LAGADIC Team.



(a) Needle insertion in the shoulder (b) Needle insertion in the leg

Figure 13. Regional anaesthesia with needle insertion and muscle contraction

5.2.2.4. Control of elastic soft robots

Participant: Frédérick Largillière.

We developed a prototype of stiffness-controlled haptic interface using a piece of silicone rubber to render different forces related to a displacement ie. different stiffnesses and an improved method of simulation using multi-rate loops to try to keep the computation real-time even with models using a large number of FEM elements. (work currently under review) We also presented the idea of a surgical robot able to virtually reconstruct its environment (ie. surrounding biological tissues) through small modifications of the algorithm used for controlling soft robots (SURGETICA 2014).

5.2.3. Image-Driven Simulation

5.2.3.1. Physics-based registration algorithms

Participant: Rosalie Plantefève.

Before targeting the augmented reality for laparoscopic operations, an important step consists in solving the initial alignment problem. Given a pre-operative image of the organ (usually a CT scan) a detailed mesh is constructed. To make the information stored in this mesh available during the operation, the mesh must be registered onto the intra operative view. However, mainly due to the pneumoperitoneum, the organ has undergone important deformation between the pre-operative images acquisition and the operation. The pre-operative shape and the intra-operative shape of the organ do not correspond. Therefore a non rigid registration is required to align the mesh and the real organ. Our registration algorithms also allowed us to work on a mean to automatically recover boundary conditions of a patient specific liver.

We created a statistical atlas [29] of the human liver to store some of the liver boundary conditions positions : the veina cava and the anchor point of the falciform ligament positions. This method was presented at MICCAI 2014. We also developed a new registration method [28] that evolves automatically from a rigid registration to a non rigid registration to solve the initial alignment problem. The method use some anatomical features of the liver such as the anchor point position of the falciform ligament. This method was presented at ISBMS 2014.

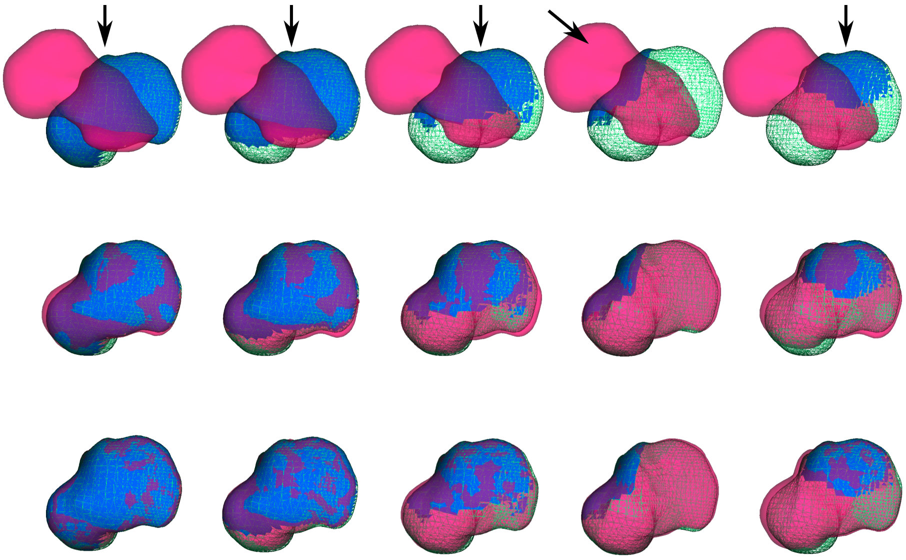


Figure 14. Results showing the initial alignment of a liver between pre-operative and intra-operative data

5.2.3.2. Augmented reality

Participant: Nazim Haouchine.

After this intra-operative registration, the augmented reality is possible. This topic is one the highlight of the year 2014. In 2014, we proposed a method for real-time augmented reality of internal liver structures during minimally invasive hepatic surgery [25]. This project is done is collaboration with the EPI MAGRIT. Vessels and tumors computed from pre-operative CT scans can be overlaid onto the laparoscopic view for surgery guidance. Compared to current methods, our method is able to locate the in-depth positions of the tumors based on partial three-dimensional liver tissue motion using a real-time biomechanical model. This model permits to properly handle the motion of internal structures even in the case of anisotropic or heterogeneous tissues, as it is the case for the liver and many anatomical structures. Experimentations conducted on phantom liver permits to measure the accuracy of the augmentation while real-time augmentation on in vivo human liver during real surgery shows the benefits of such an approach for minimally invasive surgery. Finally, a method for 3D reconstruction of elastic shapes with self-occlusion handling was also proposed.

5.2.3.3. Segmentation

Participant: Zhifan Jiang.

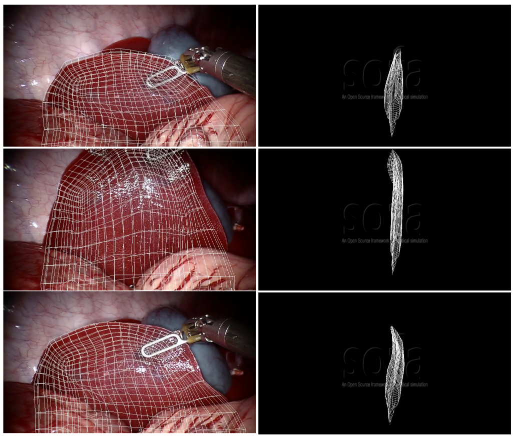


Figure 15. Augmented reality view of a liver during laparoscopic surgery

We have been working on medical image analysis in the context of the female pelvic medicine. Image-based diagnoses of pelvic floor disorders like prolapse or endometriosis rely on mechanical indicators, such as mobilities of organs and shear displacements between organs. Image data do not provide directly qualitative indicators hence analysis and diagnosis of medical are required although unfortunately subjected to surgeon expertise subjectivity. Therefore, objective information would be useful for both precise diagnoses and planning of surgical procedure. The objective is to develop numerical tools which extract quantitative information from static and cine MR images based on algorithms of detection and tracking.

We have developed numerical models not only for visualization, but also for quantitative measurements on a group of organs, such as their shapes and their relative movements. The numerical tool extracts these quantitative information (displacements and shear inter-organ) as well as the geometric shape of organs from images via Model-to-Image registration based on B-spline models. Our approach enables to identify multiple organ shapes in a single 2-dimension MR image and then to track their motion in a sequence of 2-dimension dynamic (cine) MR images for the study of the mobilities of the pelvic system. The method has been tested on healthy and pathological patient-specific data (19 patients) and the results provide valuable data to assess the shear displacement between organs and therefore making it possible to identify weakened ligaments or fascia which function differently in patients having pathologies. However, the results are to be validated by further mechanical FEM simulations. This work has been accepted in the journal STRAIN.

5.2.3.4. *MIND project*

Participants: Myriam Lekkal, Raffaella Trivisonne.

Within a feasibility study contest, we worked on Human Computer Interaction developing a new, intuitive and efficient way to interact with medical information in modern operating room. Nowadays operating rooms are progressively outfitted with computerized equipment necessary to access and manipulate a significant amount of data (i.e. medical images, patient's records, patient's vitals and physical parameters of the operating environment). This type of equipment belongs to the non-sterile section of an operating room, therefore surgeons, who are not allowed to be contaminated, cannot directly interact with it.

The idea of MIND project is to create a new device that could be used alone, such as a remote control, or easily integrated onto several locations, according to user preferences or constraints from the surgical procedure. Through this remote control, surgeons are able to access and manipulate medical information within the operating field and without leaving the instruments. For the software side the main aspects are distributed in two categories: a low level library, in charge of tasks such like handling the communications between

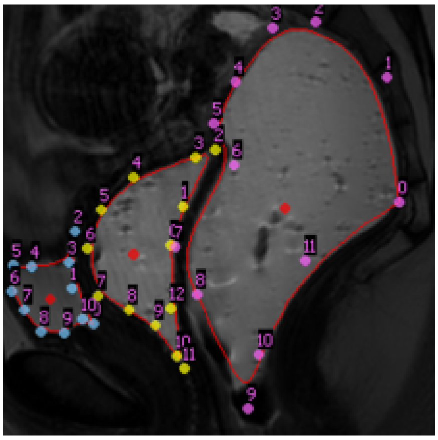
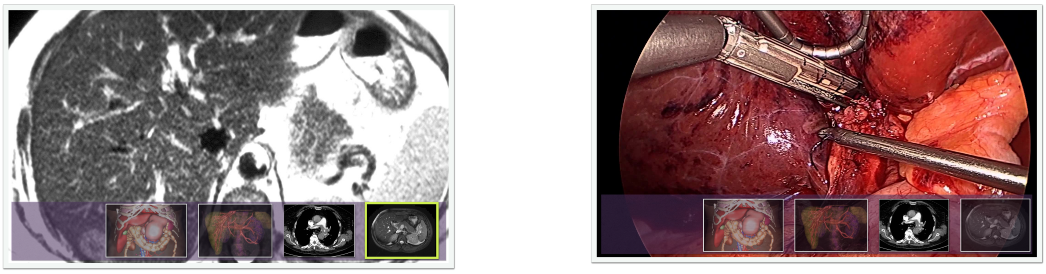


Figure 16. Contour segmentation on the pelvic system

the wireless instrument and the central computer, and a set of high-level functionalities and applications concerning the development of users GUI and new applications according to the needs of the case.

This work resulted in a patent [35] (still pending). Read more here http://mindsurgeonmouse.weebly.com/.



(a) View 1

(b) View 2

Figure 17. Example of the MIND GUI

SISTM Team

6. New Results

6.1. Highlights of the Year

A work (described below), in collaboration with M. Davis and R. Tibshirani from Standford University, has been published in the "Proceedings of the National Academy of Sciences" : [8].

Females have generally more robust immune responses than males for reasons that are not well-understood. Here we used a systems analysis to investigate these differences by analyzing the neutralizing antibody response to a trivalent inactivated seasonal influenza vaccine (TIV) and a large number of immune system components, including serum cytokines and chemokines, blood cell subset frequencies, genome-wide gene expression, and cellular responses to diverse in vitro stimuli, in 53 females and 34 males of different ages. We found elevated antibody responses to TIV and expression of inflammatory cytokines in the serum of females compared with males regardless of age. This inflammatory profile correlated with the levels of phosphorylated STAT3 proteins in monocytes but not with the serological response to the vaccine. In contrast, using a machine learning approach, we identified a cluster of genes involved in lipid biosynthesis and previously shown to be up-regulated by testosterone that correlated with poor virus-neutralizing activity in men. Moreover, men with elevated serum testosterone levels and associated gene signatures exhibited the lowest antibody responses to TIV. These results demonstrate a strong association between androgens and genes involved in lipid metabolism, suggesting that these could be important drivers of the differences in immune responses between males and females.

6.2. Analysis of purely random forests bias

In collaboration with S. Arlot, we write a research report on some theoretical results about random forests : [30].

Random forests are a very effective and commonly used statistical method, but their full theoretical analysis is still an open problem. As a first step, simplified models such as purely random forests have been introduced, in order to shed light on the good performance of random forests. In this paper, we study the approximation error (the bias) of some purely random forest models in a regression framework, focusing in particular on the influence of the number of trees in the forest. Under some regularity assumptions on the regression function, we show that the bias of an infinite forest decreases at a faster rate (with respect to the size of each tree) than a single tree. As a consequence, infinite forests attain a strictly better risk rate (with respect to the sample size) than single trees. Furthermore, our results allow to derive a minimum number of trees sufficient to reach the same rate as an infinite forest. As a by-product of our analysis, we also show a link between the bias of purely random forests and the bias of some kernel estimators.

VISAGES Project-Team

6. New Results

6.1. Highlights of the Year

Dr Camille Maumet was awarded by the French Society of Magnetic Resonance in Biology and Medicine (SFRMBM) for her PhD Thesis on analysis of neuroimaging data including images from functional Magnetic Resonance Imaging (fMRI) and Arterial Spin Labeling http://www2.warwick.ac.uk/fac/sci/wmg/idh/idhnews/?tag=Neural+Engineering.

Dr Americ Stamm was awarded by the Univ. of Rennes I foundation as the best PhD thesis in Math, Computer Sciences and Electrical Engineering. This award is dedicated for the PhDs having the highest potential for innovation and technological transfer
;ses-de-la-fondationhttps://fondation.univ-rennes1.fr/les-prix-de-thèses-de-la-fondation.

6.2. Image Computing: Detection, Segmentation, Registration and Analysis

6.2.1. A Mathematical Framework for the Registration and Analysis of Multi-Fascicle Models for Population Studies of the Brain Microstructure

Participant: Olivier Commowick.

Diffusion tensor imaging (DTI) is unable to represent the diffusion signal arising from multiple crossing fascicles and freely diffusing water molecules. Generative models of the diffusion signal, such as multi-fascicle models, overcome this limitation by providing a parametric representation for the signal contribution of each population of water molecules. These models are of great interest in population studies to characterize and compare the brain microstructural properties. Central to population studies is the construction of an atlas and the registration of all subjects to it. However, the appropriate definition of registration and atlasing methods for multi-fascicle models have proven challenging. This paper proposes [24] a mathematical framework to register and analyze multi-fascicle models. Specifically, we define novel operators to achieve interpolation, smoothing and averaging of multi-fascicle models. We also define a novel similarity metric to spatially align multi-fascicle models. Our framework enables simultaneous comparisons of different microstructural properties that are confounded in conventional DTI. The framework is validated on multi-fascicle models from 24 healthy subjects and 38 patients with tuberous sclerosis complex, 10 of whom have autism. We demonstrate the use of the multi-fascicle models registration and analysis framework in a population study of autism spectrum disorder. This work was performed in close collaboration with the Children's Hospital Boston.

6.2.2. Longitudinal Intensity Normalization in Multiple Sclerosis Patients

Participants: Yogesh Karpate, Olivier Commowick, Christian Barillot, Gilles Edan.

In recent years, there have been many Multiple Sclerosis studies using longitudinal MR images to study and characterize the MS lesion patterns. The intensity of similar anatomical tissues in MR images is often different because of the variability of the acquisition process and different scanners. We proposed [29] a novel methodology for a longitudinal lesion analysis based on intensity standardization to minimize the inter-scan intensity difference. The intensity normalization maps parameters obtained using a robust Gaussian Mixture Model (GMM) estimation not affected by the presence of MS lesions. Experimental results demonstrated that our technique accurately performs the task of intensity standardization. We show consequently how the same technique can improve the results of longitudinal MS lesion detection.

6.2.3. Simultaneous Estimation of T1, T2 and B1 Maps From Relaxometry MR Sequences

Participants: Fang Cao, Olivier Commowick, Elise Bannier, Christian Barillot.

Interest in quantitative MRI and relaxometry imaging is rapidly increasing to enable the discovery of new MRI disease imaging biomarkers. While DESPOT1 is a robust method for rapid whole-brain voxelwise mapping of the longitudinal relaxation time (T1), the approach is inherently sensitive to inaccuracies in the transmitted flip angles, defined by the B1 inhomogeneity field, which become more severe at high field strengths (e.g., 3T). We have proposed [26] a new approach for simultaneously mapping the B1 field, M0 (proton density), T1 and T2 relaxation times based on regular fast T1 and T2 relaxometry sequences. The new method is based on the intrinsic correlation between the T1 and T2 relaxometry sequences to jointly estimate all maps. It requires no additional sequence for the B1 correction. We evaluated our proposed algorithm on simulated and in-vivo data at 3T, demonstrating its improved accuracy with respect to regular separate estimation methods.

6.2.4. Quantitative Relaxation Templates for the Human Brain at 3T

Participants: Fang Cao, Olivier Commowick, Christian Barillot.

Quantitative MRI (qMRI) templates of relaxation times and proton density can be of particular interest for dedicated clinical applications such as characterizing brain tissue abnormalities, as well as general research purposes. We have developed in [27] 3D qMRI statistical templates consisting of T1, T2, T2* and ρ* maps from the human brain at 3T. The qMRI templates were built from a population of 20 normal controls, for which individual maps were estimated in a robust manner, accounting for acquisition artifacts and expected relationships between the relaxometry parameters. For validation, we fed the qMRI templates into a realistic MRI simulator to synthesize MR-weighted images, and compared these images with the real MR acquisitions. High correlation coefficients (>0.80) show that the developed qMRI templates can be used as input dataset for MRI simulation community, which may be of great interest to clinical neuroscience field.

6.2.5. Myelin Water Fraction Imaging in Multiple Sclerosis patients

Participants: Olivier Commowick, Elise Bannier, Christian Barillot.

Multi-echo T2 relaxometry is a relevant imaging method for Myelin Water Fraction (MWF) quantification in the study of multiple sclerosis (MS). However, to ensure accurate estimation, a large number of echoes are still required that can drive to very long acquisitions. In practice, 32 echo times (TE) ranging from 10 ms to 320 ms and an echo spacing (ESP) of 10 ms are used. Analysis of the decay curve of the consecutive echoes allows the estimation of the T2 spectrum. The proposed approach makes use of recent spatial regularization methods for MWF estimation from clinically compatible acquisitions (typically 11 echoes acquired within 6 minutes with TE1=ESP=8.4 ms). The algorithms were evaluated on both synthetic and clinical data, illustrating the ability to compute accurate MWF maps from a low number of echoes. The 2 methods used a priori information as well as conventional and fast algorithm (NNLS), and a cross-validation strategy. Based on simulated and clinical data results, the nlsrNNLS estimation is more accurate and less penalizing than srNNLS. This regularization provides an efficient way to circumvent an ill-posed problem aspect, in particular with a reduced number of echoes for clinically acceptable acquisition times, allowing for accurate MWF estimation. This work, performed in the master internship of Lucas Soustelle, was accepted as a conference abstract at SFRMBM 2015, and is submitted to ISMRM.

6.3. Image processing on Diffusion Weighted Magnetic Resonance Imaging

6.3.1. Fast Identification of Optimal Fascicle Configurations from Standard Clinical Diffusion MRI Using Akaike Information Criterion

Participants: Olivier Commowick, Christian Barillot.

Analytic multi-compartment models have gained a tremendous popularity in the recent literature for studying the brain white matter microstructure from diffusion MRI. This class of models require the number of compartments to be known in advance. In the white matter however, several non-collinear bundles of axons, termed fascicles, often coexist in a same voxel. Determining the optimal fascicle configuration is a model selection problem. We have proposed [30], [33] a novel approach to identify such a configuration from clinical diffusion MRI where only few diffusion images can be acquired and time is of the essence. Starting from a set of fitted models with increasing number of fascicles, we use Akaike information criterion to estimate the probability of each candidate model to be the best Kullback-Leibler model. These probabilities are then used to average the different candidate models and output an MCM with optimal fascicle configuration. This strategy is fast and can be adapted to any multi- compartment model. We illustrate its implementation with the ball-and-stick model and show that we obtain better results on single-shell low angular resolution diffusion MRI, compared to the state-of-the-art automatic relevance detection method, in a shorter processing time.

6.3.2. Tracking the Cortico-Spinal Tract as a Multi-Modal Distribution of Streamlines from Local White Matter Microstructure Models

Participant: Olivier Commowick.

We have presented [31] a pipeline to reconstruct the corticospinal tract (CST) that connects the spinal cord to the motor cortex. The proposed method combines a new white matter microstructure model coined Diffusion Directions Imaging and a new tractography algorithm based on a particle filter adapted for approximating multi-modal distributions. In this paper, we put the computation time and accuracy of our pipeline to the test in the context of the MICCAI 2014 DTI challenge, which aims to provide fast and accurate reconstructions of the CST for presurgical planning of brain tumor extraction. These two key performance metrics are expected in such a situation where time is of the essence and the quality of the data is dependent on the patient's health condition and ability to cooperate. In no more than 1.5 hours per patient, we successfully provide accurate CSTs of 2 very collaborative patients who underwent a diffusion MRI protocol that included 69 diffusion-sensitizing gradients spread over 4 different shells ranging from b = 200 to b = 3000 s/mm^2.

6.3.3. Model selection improvement with non-central chi estimation of multi-compartment models

Participant: Olivier Commowick.

Diffusion images are known to be corrupted with a non-central chi (NCC)-distributed noise. There has been a number of proposed image denoising methods that account for this particular noise distribution. However, to the best of our knowledge, no study was performed to assess the influence of the noise model in the context of diffusion model estimation. In particular, multi-compartment models are an appealing class of models to describe the white matter microstructure but require the optimal number of compartments to be known a priori. Its estimation is no easy task since more complex models will always better fit the data, which is known as over-fitting. However, MCM estimation in the literature is performed assuming a Gaussian-distributed noise. We have shown in a preliminary study [32] that using the appropriate NCC distribution for modeling the noise model reduces significantly the over-fitting, which could be helpful for unraveling model selection issues and obtaining better model parameter estimates.

6.3.4. Symmetric Block-Matching Registration for the Distortion Correction of Echo-Planar Images

Participants: Renaud Hédouin, Olivier Commowick, Elise Bannier, Christian Barillot.

We have introduced a new approach to correct geometric and intensity distortion of Echo Planar Images (EPI) from images acquired with opposite phase encoding directions. A new symmetric block-matching registration algorithm has been developed for this purpose relying on new adapted transformations between blocks and a symmetric optimization scheme to ensure an opposite symmetric transformation. We present results of our algorithm showing its ability to robustly recover EPI distortion while obtaining sharper results than the popular TOPUP algorithm. This work was performed in close collaboration with the Children's hospital in Boston.

6.4. Medical Image Computing in Brain Pathologies

6.4.1. Adaptive Dictionary Learning for Competitive Classification of Multiple Sclerosis Lesions

Participants: Hrishikesh Deshpande, Pierre Maurel, Christian Barillot.

The manual delineation of Multiple Sclerosis (MS) lesions is a challenging task pertaining to the requirement of neurological experts and high intra- and inter-observer variability. It is also time consuming because large number of Magnetic Resonance (MR) image slices are needed to obtain 3-D information. Over the last years, various models combined with supervised or unsupervised classification methods have been proposed for segmentation of MS lesions using MR images. Recently, signal modeling using sparse representations (SR) has gained tremendous attention and is an area of active research. SR allows coding data as sparse linear combinations of the elements of over-complete dictionary and has led to interesting image recognition results. In this work, we have proposed to use a sparse representation and an adaptive dictionary learning paradigm to automatically classify Multiple Sclerosis (MS) lesions from MRI. In particular, we investigate the effects of learning dictionaries specific to the lesions and individual healthy brain tissues, which include White Matter (WM), Gray Matter (GM) and Cerebrospinal Fluid (CSF). The dictionary size plays a major role in data representation but it is an even more crucial element in the case of competitive classification. We present an approach that adapts the size of the dictionary for each class, depending on the complexity of the underlying data. The proposed algorithm is evaluated on 3-D multi-channel MR images demonstrating improved classification.

6.4.2. Predictive Value of Imaging Markers at Multiple Sclerosis Disease Onset Based on Gadolinium- and USPIO- Enhanced MRI and Machine Learning

Participants: Olivier Commowick, Jean-Christophe Ferré, Elise Bannier, Gilles Edan, Christian Barillot.

A novel characterization of Clinically Isolated Syndrome (CIS) patients according to lesion patterns has been proposed in [13]. More specifically, patients are classified according to the nature of inflammatory lesions patterns. It is expected that this characterization can infer new prospective figures from the earliest imaging signs of Multiple Sclerosis (MS), since it can provide a classification of different types of lesions across patients. The method is based on a two-tiered classification. Initially, the spatio-temporal lesion patterns are classified. The discovered lesion patterns are then used to characterize groups of patients. The patient groups are validated using statistical measures and by correlations at 24-month follow-up with hypointense lesion loads. The methodology identified 3 statistically significantly different clusters of lesion patterns showing p-values smaller than 0.01. Moreover, these patterns defined at baseline correlated with chronic hypointense lesion volumes by follow-up with an R^2 score of 0.90. The proposed methodology is capable of identifying three major different lesion patterns that are heterogeneously present in patients, allowing a patient classification using only two MRI scans. This finding may lead to more accurate prognosis and thus to more suitable treatments at early stage of MS.

6.4.3. Robust detection of multiple sclerosis lesion from intensity-normalized multi-channel MRI

Participants: Yogesh Karpate, Olivier Commowick, Christian Barillot.

Multiple sclerosis (MS) is a disease with heterogeneous evolution among the patients. Better understanding of the disease will lead to improved patient-adapted therapeutic strategies. We propose a novel paradigm to detect MS lesions based on a statistical framework which consists of detection based on differences between multi-channel MRI of patients and controls. This framework fused with intensity standardization was applied to the study of MS and highlighted the great interest of quantitative MRI measurements for a better characterization of MS. Experimental results demonstrate that our technique accurately detects significant differences in lesions consequently improving the results of MS lesion detection. This work has been accepted to SPIE Medical Imaging 2015.

6.4.4. Multiple Sclerosis Lesions Recognition: One Class Learning Approach

Participants: Yogesh Karpate, Olivier Commowick, Christian Barillot, Gilles Edan.

We have developed an automatic algorithm for the detection of multiple sclerosis lesions (MSL) from multi-sequence magnetic resonance imaging (MRI). We build a probabilistic classifier that can recognize MSL as a novel class, trained only on Normal Appearing Brain Tissues (NABT). Patch based intensity information of MRI images is used to train a classifier at the voxel level. The classifier is in turn used to compute a probability characterizing the likelihood of each voxel to be a lesion. This probability is then used to identify a lesion voxel based on simple Otsu thresholding. This work has been submitted to ISBI 2015.

6.5. Vascular Imaging and Arterial Spin Labeling

6.5.1. Peripheral angiography and neurovascular imaging

Participants: Hélène Raoult, Jean-Yves Gauvrit, Elise Bannier, Pierre Maurel, Christian Barillot, Jean-Christophe Ferré.

Work-in-progress Non contrast enhanced MR angiography sequences were optimized on phantom as well as healthy volunteers and evaluated on patients presenting arterio venous malformations (AVM). High temporal resolution (70ms) images were obtained and compared to the gold standard Digital Subtraction Angiography. Results showed that Time-resolved SL MR angiographic imaging over two cardiac cycles is a reliable clinical tool for cerebral AVM characterization, yielding very good to excellent agreement with DSA. This work was published in Radioloy late 2013. This data was also post processed to obtain hemodynamics maps (time to peak, wash-in, wash out and mean transit time) and discriminate among different AVM components to relate hemodynamic patterns with rupture risk. This work was published in Stroke [23].

6.5.2. Robust perfusion maps in Arterial Spin Labeling by means of M-estimators

Participants: Pierre Maurel, Jean-Christophe Ferré, Christian Barillot.

Non-invasive measurement of Cerebral Blood Flow (CBF) is now feasible thanks to the introduction of Arterial Spin Labeling (ASL) Magnetic Resonance Imaging (MRI) techniques. To date, due to the low signal-to-noise ratio of ASL, a single acquisition (pair of control/label scans) is not sufficient to estimate perfusion reliably. Instead, the acquisition is usually repeated several times and the perfusion information is calculated by averaging across the repetitions. However, due to its zero breakdown point, the sample mean is very sensitive to outliers. We have proposed [18] to compute ASL CBF maps using Huber's M-estimator, a robust statistical function that is not overly impacted by outlier. This work was part of the PhD thesis of Camille Maumet.

6.5.3. Brain perfusion gender difference study using MRI in young adults

Participants: Léa Itmi, Pierre Maurel, Christian Barillot.

The usage of population models is becoming increasingly important in cerebral imaging, particularly in ASL. Therefore, it is important to check the limits of the models before applying them, to guarantee the reliability of the results. It is now well-known that brain perfusion changes with the age, and this effect is taken into account when evaluating brain perfusion images. But gender differences have not been well studied yet. It is known for a long time that female brain perfusion is higher than male brain perfusion, but few studies have investigated whether some regional perfusion differences exist or not. We evaluate whether, as for the age, gender differences should be taken into account when analyzing brain perfusion images. We focus on young adults subjects and studied, at the region level and the voxel level if gender differences exist and how it differs. The overall and regional differences were analyzed and then we also investigated the perfusion asymmetries in the brain (left hemisphere versus right hemisphere).

6.6. EEG and MR Imaging

6.6.1. Feasibility and specificity of simultaneous EEG and ASL MRI at 3T

Participants: Elise Bannier, Marsel Mano, Isabelle Corouge, Lorraine Perronnet, Christian Barillot.

Brain functional imaging can be performed using several approaches, including EEG, BOLD and ASL MRI. The Neurinfo platform has acquired an EEG MR compatible 64ch device over the summer to perform joint EEG and BOLD or ASL fMRI. To date, only a few studies have addressed the issue of connecting EEG signal to ASL perfusion. The aim of this study was to assess ASL-EEG at 3T in terms of safety as well as EEG and MR signal quality. The temperature measurements, specific absorption rate, and signal to noise ratio experiments have shown that ASL EEG can be safely performed using the parameters presented above. However, residual gradient artifacts in the PASL-EEG data have to be considered. Further research is needed to understand the artifact variability and to develop an appropriate correction strategy. This study is performed as part of the HEMISFER project in close collaboration with the involved teams.

6.6.2. Neurofeedback using Virtual Reality and Hybrid EEG-MRI for Brain Rehabilitation

Participants: Lorraine Perronnet, Marsel Mano, Christian Barillot.

We have conducted a thorough state-of-the-art of Neurofeedback (NF) and restorative Brain Computer Interfaces (BCI) under EEG and fMRI modality as well as of EEG-fMRI integration, with a particular focus on applications in depression and motor rehabilitation. This enabled us to build a theoretical comparison of EEG- and fMRI-NF methodology that will be helpful in designing NF protocols combining both modalities. In this perspective, we are currently designing a NF protocol based on motor imagery that will be compatible with EEG and fMRI, and running preliminary recordings of motor execution and motor imagery. Besides, we are writing a book chapter about NF and BCI that is intended to disambiguate the existing definitions and to present basic knowledge about NF principles and applications to naive readers. This is a joint work with Anatole Lécuyer team (Hybrid), in the frame of the Hemisfer project.

6.6.3. Symmetrical EEG and fMRI Imaging by Sparse Regularization

Participants: Thomas Oberlin, Pierre Maurel, Christian Barillot.

This work considers the problem of brain imaging using simultaneously recorded electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). To this end, we introduce a linear coupling model that links the electrical EEG signal to the hemodynamic response from the blood-oxygen level dependent (BOLD) signal. Both modalities are then symmetrically integrated, to achieve a high resolution in time and space while allowing some robustness against potential decoupling of the BOLD effect. The novelty of the approach consists in expressing the joint imaging problem as a linear inverse problem, which is addressed using sparse regularization. The sparsity prior naturally reflects the fact that only few areas of the brain are activated at a certain time, and it is easily implemented through efficient so-called proximal algorithms. The significance of the method and the effectiveness of the algorithms are demonstrated through numerical investigations on a simplified head model and simulated data on a realistic brain model. This is a joint work with Remi Gribonval team (Panama), in the frame of the Hemisfer project.

ANGE Project-Team

6. New Results

6.1. Highlights of the Year

In 2014, ANGE status turned from Inria team to Inria project-team. Afterwards, M. Parisot was recruited by Inria as a junior researcher.

6.2. Analysis of models in fluid mechanics

6.2.1. Well-posedness of multilayer Shallow Water-type equations

Participants: Emmanuel Audusse, Bernard Di Martino, Ethem Nayir, Yohan Penel.

The hyperbolicity of some 2-layer Shallow Water equations had been proven in [26], [23], there are many open theoretical investigations to lead about these systems. In particular, E. Nayir proved the local well-posedness of the model derived in [23] for periodic boundary conditions. Next steps will consist in extending this preliminary result to the whole space and proving the global existence of strong solutions. The existence of weak solutions will be studied from B. Di Martino's work. The hyperbolicity for N layers must also be investigated.

As for numerical aspects, the use of FRESHKISS3D will provide qualitative assessments for modelling issues (viscous tensor, source terms, variable density, interfacial velocities). It will also yield comparisons with theoretical results, in particular when the number of layers goes to infinity.

6.2.2. Non-hydrostatic models

Participants: Dena Kazerani, Jacques Sainte-Marie, Nicolas Seguin.

Together with Corentin Audiard from Univ. Pierre et Marie Curie, we investigated the structure of general non hydrostatic models for shallow water flows. This includes the Green–Naghdi equations and the model proposed by Bristeau *et al.* in [13]. D. Kazerani proved that such systems possess a symmetric structure based on the existence of an energy. The main difference with the well-known hyperbolic case is due to the presence of differential operators instead of matrices.

6.3. Modelling of complex flows

6.3.1. Dynamics of sedimentary river beds with stochastic fluctuations

Participants: Emmanuel Audusse, Philippe Ung.

We studied in [9] the behaviour of the solution of the Saint-Venant–Exner equations when a stochastic term is introduced in the model through the sediment flux. A first investigation was done considering periodic boundary conditions and the next part of this study is devoted to the case when physical ones are imposed. Our goal is to investigate the possibility to bring out a characteristic long time behaviour and to establish a relation between the injected noise and the physical parameters involved in the model. This work was achieved in collaboration with Sébastien Boyaval from Lab. Hydraulique Saint-Venant.

6.3.2. Non-hydrostatic effects

Participants: Nora Aïssiouene, Marie-Odile Bristeau, Edwige Godlewski, Dena Kazerani, Anne Mangeney, Jacques Sainte-Marie, Nicolas Seguin.

The objective is to derive a model corresponding to a depth averaged version of the incompressible Euler equations with free surface and to develop a robust numerical method for the resolution of the model.

Concerning the modelling aspect, a non-hydrostatic shallow water-type model approximating the incompressible Euler and Navier-Stokes sytems with free surface was developped and published in [13]. The closure relations are obtained by a minimal energy constraint instead of an asymptotic expansion. The model slightly differs from the well-known Green-Naghdi model and is confronted with stationary and analytical solutions of the Euler system corresponding to rotational flows.

The numerical approximation relies on a projection-correction type scheme. The hyperbolic part of the system is approximated using a kinetic finite volume solver and the correction step implies to solve an elliptic problem involving the non-hydrostatic part of the pressure.

In one dimension, the resolution of the incompressibility problem leads to solve a mixed problem where the pressure and the velocity are defined in compatible approximation spaces. This step uses a variationnal formulation of the shallow water version of the incompressibility condition.

This numerical scheme satisfies classical properties (positivity, well-balancing and consistency) and a discrete entropy inequality. Several numerical experiments are performed to confirm the relevance of our approach.

This approach will allow us to extend the numerical method in higher dimensions and to treat particular difficult cases occuring in specific geophysical situations (dry/wet interfaces).

6.3.3. Plasticity in Shallow Water equations

Participant: Nicolas Seguin.

In collaboration with Bruno Després and Clément Mifsud from Univ. Pierre et Marie Curie, we proposed in [20] a new definition of solutions for hyperbolic Friedrichs' systems in bounded domains, which follows the idea of Lions' dissipative solutions and Otto's boundary formulation for conservation laws. We proved in the classical settings existence and uniqueness. The goal of this project is to be able to incorporate nonlinear effects of plasticity in models of elasticity or overflowing in channels for shallow water flows, by adding entropy compatible constraints.

6.3.4. Management of marine energies

Participants: Cindy Guichard, Martin Parisot, Jacques Sainte-Marie, Julien Salomon.

The purpose of this project is to model floating devices (like buoys) in the context of recovering energy from water resources (seas and oceans). If the free surface flow can be handled by means of the Saint-Venant equations, the area under the buoys requires a different modelling (for example equivalence with springs) as the surface is constrained. The Archimedes' principle is also involved. Some preliminary numerical results were obtained thanks to the FRESHKISS3D code.

To go further, the optimisation of the overall process is also under consideration. Indeed, to maximise the amount of recovered energy, the bathymetry, the shape of the buoy, the number of buoys are critical parameters which must be modelled in view of industrial applications. Optimal control methods are applied to determine the best configuration depending on the devices: optimisation of the kinetic energy for water-turbines or of the potential energy for buoys.

6.4. Accurate simulations of fluid flows

6.4.1. A numerical scheme for the Saint-Venant–Exner equations

Participants: Emmanuel Audusse, Philippe Ung.

After having established a Godunov-type method based on the design of a three-wave Approximate Riemann Solver for the Saint-Venant equations [10], we extended this approach to the Saint-Venant–Exner equations for modelling the sediment transport. The coupled aspect between the hydraulic and the morphodynamic parts is only located on the evaluation of the wave velocities. Under this assumption, the proposed scheme can be interpreted as a hybrid method between the splitting and non-splitting methods and it also raises the issue of the choice between the two previous approaches.

These results were proven in collaboration with Christophe Chalons from Univ. Versailles–Saint-Quentin.

6.4.2. *Simulations of fluid/particules interactions*

Participant: Nicolas Seguin.

In collaboration with Nina Aguillon and Frédéric Lagoutière from Univ. Paris-Sud, we proved in [7] the convergence of finite volume schemes for a simplified model of fluid-particle interaction. The mesh follows the particle which appears in the model as a pointwise contribution. The numerical scheme is based on local well-balanced fluxes, which permits to obtain compactness and convergence.

6.4.3. *Hydrostatic reconstruction*

Participants: Emmanuel Audusse, Marie-Odile Bristeau, Jacques Sainte-Marie.

The hydrostatic reconstruction is a general and efficient method to handle source terms that uses an arbitrary solver for the homogeneous problem and leads to a consistent, well-balanced, positive scheme satisfying a semi-discrete entropy inequality.

In [8], we proved with Francois Bouchut from Univ. Marne-la-Vallée that the hydrostatic reconstruction coupled to the classical kinetic solver satisfies a fully discrete entropy inequality which involves an error term but the latter goes to zero strongly with the mesh size.

6.4.4. *A numerical scheme for multilayer shallow-water model for all Froude regimes*

Participant: Martin Parisot.

The aim of this work in collaboration with Jean-Paul Vila from INSA/IMT is to propose an efficient numerical resolution to simulate stratified non-miscible fluids. The strategy should be consistent for all regime especially with the so-called low-Froude regime particularly relevant for applications. The proposed scheme is entropy-satisfying, well-balanced and asymptotic preserving. In addition the stability of the scheme is ensured for large time scale. More precisely, it does not depend on the gravity waves, which are very restrictive for the targeted applications, such as oceanology and meteorology. Further work using the strategy for sustainable energies is in progress.

6.4.5. *Adaptation of the Godunov scheme to the low Froude regime*

Participants: Emmanuel Audusse, Do Minh Hieu, Yohan Penel.

Standard numerical schemes designed for the simulation of fluid flows are known to fail when the Mach number becomes too small. Similar behaviours are observed for geophysical flows when the Froude number decreases. Do Minh Hieu is interested in the numerical simulation of the Shallow Water equations including some Coriolis forces. He investigated several corrections of the standard Godunov schemes in 1D to preserve the kernel of spatial operators involved in the aformentioned equations and blamed for being responsible of the loss of accuracy. He now intends to perform the same analysis in 2D under the supervision of E. Audusse, S. Dellacherie (from CEA), P. Omnès (from CEA) and Y. Penel.

6.5. Software development and assessments

6.5.1. *Improvements in the FRESHKISS3D code*

Participants: Marie-Odile Bristeau, David Froger, Raouf Hamouda, Jacques Sainte-Marie.

Several tasks have been achieved in the FRESHKISS3D software:

- FreshKiss3D has been improved to take into account the second order in space for the 3D cases.
- The solver now includes the second order in time.
- The numerical validation using 3D numerical analytical solutions has been achieved.
- Numerous simulations have been driven by industrial contracts:
  - Simulations of fluid hydrodynamics in lagoons for optimizing the geometric field to ensure a high level of agitation for a low energy consumption (SAUR)
  - Simulations of fluid hydrodynamics in lagoons showing the vertical distribution of velocity and how to use it for optimizing micro-algae production (Salinalgue)
- Tsunamis simulations leading to the module TsunaMaths, web interface showing some historical tsunamis.
- Geometric implementations of FRESHKISS3D have been improved.
- Unit tests are being made automatically as the source code is modified.
- A user interface has been created using Python.
- The parallelization of FRESHKISS3D with MPI is under development.

CASTOR Project-Team

6. New Results

6.1. Physics

6.1.1. Physical studies

6.1.1.1. Parallel Kelvin-Helmholtz instability

Participants: Hervé Guillard, Marco Bilanceri, Céline Colin [CEA], Philippe Ghendrih [CEA], Giorgio Giorgiani, Boniface Nkonga, Frédéric Schwander [M2P2, AMU], Eric Serre [M2P2, AMU], Patrick Tamain [CEA].

In the scrape-off layer (SOL) of tokamaks, the flow acceleration due to the presence of limiter or divertor plates rises the plasma velocity in a sonic regime. These high velocities imply the presence of a strong shear between the SOL and the core of the plasma that can possibly trigger some parallel shear flow instability. The existence of these instabilities, denoted as parallel Kelvin-Helmholtz instability in some works have been investigated theoretically in [51] using a minimal model of electrostatic turbulence composed of a mass density and parallel velocity equations. This work showed that the edge plasma around limiters might indeed be unstable to this type of parallel shear flow instabilities. In this work, begun in 2013, we have performed large scale 3D simulations using the PlaTo platform of the same simple mathematical model to investigate this question. The numerical results confirm that in agreement with the theoretical expectations as well as with other numerical methods, the sheared flows in the SOL are subject to parallel Kelvin-Helmholtz instabilities. However, the growth rate of these instabilities is low and these computations require both a sufficient spatial resolution and a long simulation time. This makes the simulation of parallel Kelvin-Helmholtz instabilities a demanding benchmark but it also allows us to validate the parallel implementation of the PlaTo platform up to up to $\mathcal{O}(1000)$ CPU [14].

6.2. Numerical developments

6.2.1. Numerical developments

6.2.1.1. Conformal hexahedral mesh coarsening by agglomeration

Participants: Hervé Guillard, Pierre Cargemel, Youssef Mesri [IFPEN].

This work has been realized in the framework of a PhD contract with IFPEN that aims to produce a coarsening software for hex-dominant meshes. Reservoir simulation involves to compute dynamic flow of different phases in a porous medium. The initial state of the reservoir is usually pre-computed via geo-statistics methods extrapolating measures of the terrain. Therefore, the input of reservoir simulation is given as a very fine mesh containing heterogeneous data and numerical simulation on this fine mesh is usually non-practical. This work is therefore devoted to the study of an agglomeration strategy, to dynamically coarsen this fine hex-dominant mesh. The adaptivity may be driven by physics and/or geometric estimators. Ideally, the coarsening should be applied locally in low gradient regions, whereas high gradient regions keep the fine mesh. This work has been presented in the 23rd International meshing Roundtable [26]. The planned sequel of this work consists to use the notion of Central Voronoi Tesselation (CVT) to treat the regions where the mesh is not structured and to apply this strategy in different physical contexts from plasma physics to petroleum engineering.

6.2.1.2. Mapped Fourier Methods for stiff problems in toroidal geometry

Participant: Hervé Guillard.

Due to the particular geometry of tokamaks, a lot of numerical codes developed for their numerical simulations use Fourier methods. Fourier spectral or pseudo-spectral methods are extremely efficient for periodic problems. However this efficiency is lost if the solutions have zones of rapid variations or internal layers. For these cases, a large number of Fourier modes are required and this makes the Fourier method unpractical in many cases. This work investigates the use of mapped Fourier method as a way to circumvent this problem. Mapped Fourier method uses instead of the usual Fourier interpolant the composition of the Fourier interpolant with a mapping in such a way that in the computational space, the functions to represent are not stiff. This work gives some examples of the usefulness of this method and apply it to a simple model of pellet injection in tokamaks as an example of its potential interest for complex multi dimensional problem [34].

6.2.1.3. Multislope MUSCL method for general unstructured meshes

Participants: Hervé Guillard, Clément Le touze [ONERA], Angelo Murrone [ONERA].

To increase the accuracy in finite volume method, the concept of MUSCL reconstruction has been introduced in the pioneering work of van Leer in the 70'. This technique is still one of the most efficient to deal with the existence of discontinuous solutions in numerical simulations. In the MUSCL technique, a discontinuous linear approximation of the solution is reconstructed on each control volume. The main approximation problem of this method is therefore to reconstruct the slope of the solution.

The multislope concept has been recently introduced in the literature to deal with MUSCL reconstructions on triangular and tetrahedral unstructured meshes in the finite volume cell-centered context. Dedicated scalar slopes are used to compute the interpolations on each face of a given element, in opposition to the monoslope methods in which a unique limited gradient is used. The multislope approach reveals less expensive and potentially more accurate than the classical gradient techniques. Besides, it may also help the robustness when dealing with hyperbolic systems involving complex solutions, with large discontinuities and high density ratios. In this work, we have designed a generalized multislope MUSCL method for cell-centered finite volume discretizations. The method is freed from constraints on the mesh topology, thereby operating on completely general unstructured meshes. Moreover optimal second-order accuracy is reached at the faces centroids and the scheme is L^∞ stable. Special attention has also been paid to equip the reconstruction procedure with well-adapted dedicated limiters, potentially CFL-dependent. We have shown in [18] the ability of the method to deal with completely general meshes, while exhibiting second-order accuracy.

6.2.1.4. Development of a two temperature model

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

A two temperature (ions - electrons) model for non-magnetized plasma has been designed. The numerical scheme is a finite volume method with an approximate Riemann solver using the total energy equation and the electron entropy as main variables. This Riemann solver has been validated against standard shock tube problems and incorporated in the PlaTo platform. The solver has been implementated in toroidal geometry and tested successfully on realistic particular flows encountered in this context. The development of a reduced MHD model based on this two temperature scheme is currently studied.

6.2.1.5. Entropy Preserving Schemes for Conservation Laws

Participants: Christophe Berthon [University of Nantes], Bruno Dubroca [CEA/DAM/CESTA and University of Bordeaux 1], Afeintou Sangam.

A relaxation-type scheme has been proposed to approximate weak solutions of Ten-Moments equations with source terms [*2*]. These equations model compressible anisotropic flows. Following the technique introduced in [44], the proposed scheme is proved to be entropy preserving.

6.2.1.6. Eurofusion WPCD: Free boundary equilibrium code and control

Participants: Cédric Boulbe, Blaise Faugeras, Jean François Artaud [IRFM CEA Cadarache], Vincent Basiuk [IRFM CEA Cadarache], Emiliano Fable [Max-Planck-Institut für Plasmaphysik, Garching], Philippe Huyn [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Jakub Urban [IPP, Academy of Sciences of the Czech Republic, Prague].

Our team is involved in the integrated modelling WPCD (Work Package Code Development) Eurofusion. This project is the continuation of the EFDA-ITM project. The goal of WPCD is to provide a european tool for tokamak simulations. Different physical codes can be coupled using Kepler environment. Machine description and physical data have been described using CPO (Consistent Physical Objet) which are used as standardized inputs and outputs for the codes.

In this project, we participate in the coupling of a free boundary equilibrium solver, the European Transport Solver (ETS) and a plasma shape and position controller. The workflow coupling TCV hybrid Simulink controller and Cedres++ using PF circuit connections has been finalized and tested on the TCV tokamak.
A new workflow coupling Cedres++ with ETS and the TCV controller has been developed and is being tested on a TCV test case. This workflow is an evolution of the coupling CEDRES++ - ETS described in [12].
A successful benchmark between the three free boundary equilibrium codes CEDRES++([15] [47]) , FREEBIE ([43]), and SPIDER ([49]) has been done on static test cases. This activity will be continued to compare the time dependent versions of the three codes.

6.2.1.7. Optimal control for scenario optimization of discharges in tokamaks

Participants: Jacques Blum, Holger Heumann.

In this project we aim for an automatic determination of optimal voltage evolutions via an optimal control formulation based on a system of partial differential equations that describes the evolution of plasma equilibrium in a tokamak. Optimal voltage evolutions are the one that ensure that the evolution of the plasma runs through predescribed, user-defined states, defined e.g. as desired evolution of shape or position. The system of partial differential equations describing the evolution of the plasma is non-linear and we use a finite element formulation together with implicit time stepping for the discretization [15]. With this approach we end up with a large but finite dimensional optimization problem with non-linear constraints. We are using SQP (sequential quadratic programming), known to be one of the fastest methods of such problems, to solve the finite-dimensional optimization problem. The performance of SQP relies on accurate derivatives of the objective function and the constraints. The derivatives related to the free boundary, derived and implements during H. Heumann's PostDoc 2011/2012 for a static optimal control problem, appeared here again and are one of the important building blocks for treating the transient case. Both in CEDRES++ and FEEQ.M we have now the capability to solve first test cases to define optimal voltage evolutions. In contrast to the static case, where the linear algebraic systems in the SQP iteration remain reasonable small, the solution of the corresponding linear system in the transient case becomes very time-consuming, which somehow limits the applicability. We are testing variants of SQP, such as BFGS-like updates for the reduced Hessian, to see whether we could speed up and improve robustness of our calculations. Fast iterative solver for large sparse linear systems is another option that we started to investigate. Fast iterative solver for linear system in transient optimal control problems governed by partial differential equations is a very active area of research and we hope to benefit from the latest developments.

6.2.1.8. Boundary reconstruction for the WEST tokamak with VacTH

Participants: Jacques Blum, Sylvain Bremond [IRFM CEA Cadarache], Cédric Boulbe, Blaise Faugeras, Holger Heumann, Philippe Moreau [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Remy Nouailletas [IRFM CEA Cadarache], François Saint Laurent [IRFM CEA Cadarache].

This work is under progress in collaboration with the CEA. The control of the plasma in the future WEST tokamak requires the identification of its boundary in real time during a pulse. The code VacTH under development in the team enables such an identification. Several numerical developments and experiments have been conducted in order to prepare the control of the plasma in the WEST tokamak. The equilibrium code CEDRES++, also developed in the team, is used to simulate a real plasma and to generate synthetic magnetic measurements from which the plasma boundary is reconstructed using the code VacTH. A control algorithm developed by the colleagues from the CEA then uses this knowledge of the plasma shape to adapt the currents flowing in the poloidal field coils in order to achieve a desired evolution of the plasma.

6.2.1.9. Equilibrium reconstruction for ASDEX UpGrade (AUG) with Vacth-Equinox

Participants: Blaise Faugeras, Rui Cohelo [IPFN, IST, Lisbonne], Patrick Mccarthy [National University of Ireland University College Cork].

Within the framework of the WPCD EUROFUSION the code VacTH-Equinox has been adapted to enable equilibrium reconstruction for AUG. The identification of the current density pedestal required the development of a specific regularization scheme allowing weaker regularization close to the plasma boundary and stronger close to the magnetic axis.

6.2.1.10. Taylor-Galerkin stabilized Finite Element

Participants: José Costa, Boniface Nkonga.

The theoretical part of Taylor-Galerkin/Variational multi-scales (TG/VMS) strategy applied to MHD and reduced MHD modeling has been achieved last year. The final method amounts to add in the finite element formulation, a self-adjoint operator associated to the most critical hyperbolic component of the system to be solved. The design of the critical contours and the identification of associated waves to be stabilized is problem dependent and related to the Jacobian matrix. We have focused this year on the validations of this strategy and the improvement of the linearization used for stabilization. For application to plasma configurations with X-point, we need to reconsider the consistency with equilibrium and the Bohm boundary conditions on open flux walls.

6.2.1.11. Toward full MHD numerical modeling with $\mathcal{C}^1$ FE

Participants: José Costa, Giorgio Giorgiani, Hervé Guillard, Boniface Nkonga.

In this context the single fluid full MHD model is considered and the divergence free constraint on the magnetic field is achieved by introduction of a potential vector. The use of the potential vector has the additional advantage that the toroidal component is the magnetic flux of the Grad-Shafranov equilibrium. However, using the potential vector as variable introduces higher order derivatives in the system and classical C0 finite elements cannot be directly applied. This is why our finite element strategies use shape/test functions whose derivatives have global continuity in space (smooth finite elements). The global approach uses cross product shape/test functions between poloidal(2D) and toroidal(1D). In the 2D poloidal plane, discretization uses either quadrangular or triangular elements.

This year we have focused on the numerical analysis associated to the full MHD discretization in configurations with open flux surfaces. In order to derive efficient strategies for the full MHD in the potential vector formulation, the Gauge condition on the potential vector and the boundary conditions have been enforced by penalizations. For the Gauge condition it gives rise to element contributions but also boundary integrals that should be computed on curved surfaces that sometime fitted the magnetic surfaces. Equations are formulated in semi-conservative form such as to apply integration by part. Therefore, boundary conditions can be viewed as evolution of fluxes or variables. Integral formulation on the boundary is very useful for higher order finite elements and also for easier treatment of corners. Indeed in this context the boundary conditions are edges/surfaces oriented and boundary corners are driven by the neighborhood edges penalizations. This strategy is the one that will be used for future developments.

2D Quadrangular Cubic Bezier Finite Elements: This finite element is used for a while for reduced MHD models in the software Jorek. Reduced MHD is used to project the momentum equation in a space orthogonal to the equilibrium. When full MHD models are used, the momentum equation needs to be projected in the equilibrium space and this projection should be consistent with the Grag-Shafranov equilibrium that is used to compute the initial state. This has been achieved by a proper computation of the JxB contribution in the momentum equation, taking into account the poloidal variation of the toroidal component of the magnetic field. Detailed analysis has been performed. The next year will be devoted to implementations and numerical validations.

2D Triangular Powel-Sabin Finite Elements: In order to avoid some mesh singularities when using quadrangular meshes for complex geometries and flux surfaces shapes, triangular elements are a possible option. It is not so easy to derive smooth finite element on triangle with reduced number of degree of freedom. The Bell reduced-quintic finite elements we have considered in the previous years have too much unknowns (6 per vertex). Powell-Sabin splines are piecewise quadratic polynomials with a global $\mathcal{C}^1$-continuity and 3 unknowns per vertex, they have a local support, they form a convex partition of unity, they are stable, and they have a geometrically intuitive interpretation involving control triangles. Construction of the Powel-Sabin splines needs some geometrical tools that have been developed: Minimum area enclosing triangle of a set of control points (sequential and parallel). This construction is applied to each vertex of the triangular mesh and used to derive the local shape/test functions. These Powel-Sabin splines have been used successfully in the area of computer aided geometric design for the modeling and fitting of surfaces. We have used the Powell-Sabin (PS) splines for the approximation of elliptic partial differential equations (including Grad-Shafranov) in a rectangular domain. In this context have recovered the optimal rate of convergence (order 3). Preliminary result has been obtained for hyperbolic isothermal 2D Euler equations with TG/VMS stabilization. Our aim in the coming years is to apply these PS splines to full MHD in a toroidal geometry.

6.2.1.12. Genuinely multidimensional Riemann Solver

Participants: Jeaniffer Vides, Boniface Nkonga.

Multidimensional Riemann solvers were pioneered by Abgrall. Abgrall, Maire, Nkonga, Després and Loubere have extensively developed them especially as node-solvers for Lagrangian hydrodynamics. Another strain of work comes from explorations by Wendroff and Balsara who took a space-time approach. In this work, the resolved state is obtained via space-time integration over a wave model, just as was done by Wendroff and Balsara. However, an algebraic approach is used for the development of the fluxes. It is, therefore, shown that the multidimensional fluxes can be obtained by application of jump conditions at the boundaries of the wave model. The problem is of course over determined with the result that the shock jump conditions are only satisfied approximately in a least squares sense. Even so, this work gives us new perspective on multidimensional Riemann solvers. The litteral satisfaction of the shock jump conditions (up to least squares approximation) makes it easier to understand multidimensional Riemann solvers as a natural extension of the one-dimensional Riemann solvers. Contributions have also been made on the development of a minimalist wave model, which might help in reducing dissipation. Further innovations are reported on the assembling of fluxes based on the structure of the wave model, and those innovations are potentially useful. For MHD the CT approach consists of constraining the transport of magnetic field so that the divergence is always kept zero. The method relies on exploiting the dualism between the flux components and the electric field. Since the electric field is needed at the edges of the mesh, the multidimensional Riemann solver can also provide the electric field. By running an extensive set of simulations, it is shown that the multidimensional Riemann solver is robust and can be used to obtain divergence-free formulations for MHD that perform well on several stringent calculations. Future work will improve this strategy by enriching the description of the strongly interaction of waves.

6.2.1.13. Multi scales approximations of "Shallow water" flows

Participants: Jeaniffer Vides, Boniface Nkonga.

The terminology "Shallow water" is used to characterize thin flows on curved surfaces. It is customary for this type of flows to use the incompressible Navier-Stokes equations to asymptotically derive reduced models for the evolution of the depth integrated speed and the thickness of the flow. Reduced model are mainly hyperbolic and finite volume method are often used for their numerical approximation. Approximations strategies are generally structured as follow:

- Construction of a global coordinate system associated with an assumed analytical surface.
- Reduction of the model relatively to the global coordinate system
- Approximation of the surface by a finite number of elements.
- Approximation of the reduced model using the discrete surface.

In the context of real applications, it is presumptuous to expect an analytical formulation of the surface. From the data provided by observation satellites, we can usually extract a discrete description of the surfaces that drives thin flow. Therefore, it is more practical to use the discrete description as the starting point of the resolution strategy. This is the angle of approach that we have considered. We locally define two mesh scales: the element scale and the cell scale. The discrete mapping and the reduced model are defined at the element scale and the average values that evolve in time are defined at the cell scale. First applications have been successfully performed. We will now continue your investigations and include relevant physics at each scale, including sheared flows. We will also examine the use of multi dimensional Riemann solver in this context.

6.2.1.14. Computational Magnetohydrodynamics with Discrete Differential Forms

Participants: Holger Heumann, Ralf Hiptmair [SAM, ETH Zürich, Switzerland], Cecilia Pagliantini [SAM, ETH Zürich, Switzerland].

Differential forms, or equivalently exterior calculus, are a natural framework for electromagnetics; not only for a better understanding of the theoretical foundation, but also for the development of numerical methods. Keywords are the Hodge decomposition or the de Rham complex that are at the bottom of recent development of efficient multigrid methods or stable mixed finite element methods. Thinking in terms of such co-ordinate free differential forms offers considerable benefits as regards to the construction of structure preserving spatial discretizations.

In the present project, we aim at developing a new approach for the numerical treatment of resistive magnetohydrodynamics where a Galerkin discretization of the electromagnetic part based on finite element exterior calculus (FEEC) will be coupled to advanced finite volume methods for the approximation of the balance laws for the fluid.

The latest developments involved the extension and analysis of the stablized Galerkin schemes for advection of differential forms introduced in [48] to the case of time-dependent and non-regular flow fields.

6.2.1.15. Hamilton-Jacobi Formulation for Vlasov-Poisson

Participants: Holger Heumann, Eric Sonnendrücker [IPP, Max-Planck-Institute Garching, Germany], Philip J. Morrison [Institute for Fusion Studies, Austin, USA].

The phase space mapping induced by the solution of the Vlasov-Poisson problem is a symplectic mapping (or canonical transformation in physics literature) solving Hamilton's equations. In this project we are developing numerical methods that are based on this formulation. We derived and implemented new finite difference schemes for the corresponding Hamilton-Jacobi equation, that circumvent the projection of the distribution function inherent in Lagrangian methods. First numerical results for standard test problems show the ability of increased resolution of fine-scale effects.

6.2.1.16. Entropy viscosity technique

Participants: Richard Pasquetti, Jean-Luc Guermond [Texas A & M University], Boyan Popov [Texas A & M University].

The entropy viscosity technique allows to address hyperbolic equations by introducing a strongly non linear viscous term where needed, especially at shocks. The basic idea is to set up a viscosity from the residual of the entropy inequality together with a $O(h)$ upper bound proportional to the local wave speed. In view of addressing situations where vaccum may appear in the tokamak, we have considered the shallow water equations with topography and in situations where dry-wet transitions occur. Using a RK scheme in time and a spectral element method (SEM) in space, we have proposed a variant of the entropy technique, that mainly consists of using the viscosity upper bound in the dry parts, to obtain satisfactory results. This work was presented in [30], [22] and a publication is under review.

6.2.1.17. Bohm boundary conditions

Participants: Sébastian Minjeaud, Richard Pasquetti.

In the frame of the ANR project ESPOIR, our partners have proposed a penalty method to enforce the Bohm criterion (Mach number greater than one at the tokamak plates). This approach has been justified by considering a “minimal transport model” that consists of a 1-dimensional non linear hyperbolic system of two equations, that govern the evolutions of the density and velocity. The approach and further developments are described in three recent papers published in the journal of computational physics. Considering the same hyperbolic system, we have proposed a direct way to enforce the Bohm criterion in the frame of an explicit time marching. Using a SVV stabilized SEM it is then possible to resolve the same problem with spectral accuracy. This paper is now in press and will be published as a JCP note.

6.2.1.18. A numerical scheme for fluid-particules flows

Participants: Florent Berthelin, Thierry Goudon, Sebastian Minjeaud.

We propose a numerical scheme for the simulation of fluid-particles flows with two incompressible phases. The numerical strategy is based on a finite volume discretization on staggered grids, with a flavor of kinetic schemes in the definition of the numerical fluxes. We particularly pay attention to the difficulties related to the volume conservation constraint and to the presence of a close-packing term which imposes a threshold on the volume fraction of the disperse phase. We are able to identify stability conditions on the time step to preserve this threshold and the energy dissipation of the original model. The numerical scheme is validated with the simulation of sedimentation flows.

6.2.1.19. Identification and forecast of ionospheric disturbances

Participants: Didier Auroux, Sebastian Minjeaud.

In the framework of ANR IODISSEE, in order to identify (and forecast) ionospheric disturbances leading to temporary losses of satellite-to-earth communications (GPS, Galileo), we used Striation software for data assimilation. We obtained the adjoint code thanks to automatic differentiation (Tapenade software from Inria). As the data from Demeter satellite were not available, we extracted synthetic data from a generic model run, and we tried to identified some physical parameters (electronic density, atomic mass, number of particles) of the initial condition from the observations. For a small physical time scale (approximately 1 hour), the identification works very well, and it is possible to retrieve the initial condition from a sparse and noisy observations, allowing us to forecast the evolution of the ionospheric plasma - and then to forecast the disturbances and plasma bubbles that trap GPS and Galileo signals. For longer physical time windows (5 to 10 hours), the identification does not work anymore. We plan to work with real data, if possible, and also with a more complex model (for instance Dynamo software).

CLIME Project-Team

6. New Results

6.1. Highlights of the Year

BEST PAPER AWARD :

[20] **VISAPP - International Conference on Computer Vision Theory and Applications**. D. BÉRÉZIAT, I. HERLIN.

6.2. State estimation: analysis and forecast

One major objective of Clime is the conception of new methods of data assimilation in geophysical sciences. Clime is active on several challenging aspects: non-Gaussian assumptions, multiscale assimilation, minimax filtering, etc.

6.2.1. An iterative ensemble Kalman smoother

Participants: Marc Bocquet, Pavel Sakov [BOM, Australia].

The iterative ensemble Kalman filter (IEnKF) was proposed for improving the performance of the ensemble Kalman filter on strongly nonlinear geophysical models. IEnKF can be used as a lag-one smoother and extended to a fixed-lag smoother: the iterative ensemble Kalman smoother (IEnKS). IEnKS is an ensemble variational method. It does not require the use of the tangent of the evolution and observation models, nor the adjoint of these models: the required sensitivities (gradient and Hessian) are computed from the ensemble. Looking for the optimal performance, we consider a quasi-static algorithm, out of the many possible extensions. IEnKS was explored on the Lorenz'95 model and on a 2D turbulence model. As a logical extension of IEnKF, IEnKS significantly outperforms the standard Kalman filters and smoothers in strongly nonlinear regimes. In mildly nonlinear regimes (typically synoptic scale meteorology), its filtering performance is marginally but clearly better than the standard ensemble Kalman filter, and it keeps improving as the length of the temporal data assimilation window is increased. For long windows, its smoothing performance very significantly outranks the standard smoothers, which is believed to stem from the variational but flow-dependent nature of the algorithm. For very long windows, the use of a multiple data assimilation variant of the scheme, where observations are assimilated several times, is advocated. This paves the way for finer re-analysis freed from the static prior assumption of 4D-Var, but also partially freed from the Gaussian assumptions that usually impede standard ensemble Kalman filtering and smoothing.

6.2.2. Modeling and assimilation of lidar signals

Participants: Yiguo Wang [CEREA], Karine Sartelet [CEREA], Marc Bocquet, Patrick Chazette [LSCE, France].

In this study, we investigate the ability of the chemistry transport model (CTM) Polair3D of the air quality platform Polyphemus to simulate lidar backscattered profiles from model aerosol concentration outputs. This investigation is an important pre-processing stage of data assimilation (validation of the observation operator). To do so, simulated lidar signals are compared to hourly lidar observations performed during the MEGAPOLI (Megacities: Emissions, urban, regional and Global Atmospheric POLlution and climate effects, and Integrated tools for assessment and mitigation) summer experiment in July 2009, when a ground-based mobile lidar was deployed around Paris on-board a van. The comparison is performed for six days (1, 4, 16, 21, 26 and 29 July 2009), corresponding to different levels of pollution and different atmospheric conditions. Overall, Polyphemus reproduces well the vertical distribution of lidar signals and their temporal variability, especially for 1, 16, 26 and 29 July 2009. Discrepancies on 4 and 21 July 2009 are due to high-altitude aerosol layers, which are not well modeled. In the second part of this study, two new algorithms for assimilating lidar observations based on the optimal interpolation method are presented. One algorithm analyses PM_{10} (particulate matter with diameter less than 10 μm) concentrations. Another analyses $\mathrm{PM}_{2.5}$ (particulate matter with diameter less than 2.5 μm) and $\mathrm{PM}_{2.5-10}$ (particulate matter with a diameter higher than 2.5 μm and lower than 10 μm) concentrations separately. The aerosol simulations without and with lidar Data Assimilation (DA) are evaluated using the Airparif (a regional operational network in charge of air quality survey around the Paris area) database to demonstrate the feasibility and usefulness of assimilating lidar profiles for aerosol forecasts. The evaluation shows that lidar DA is more efficient at correcting PM_{10} than $\mathrm{PM}_{2.5}$, probably because $\mathrm{PM}_{2.5}$ is better modeled than PM_{10}. Furthermore, the algorithm which analyzes both $\mathrm{PM}_{2.5}$ and $\mathrm{PM}_{2.5-10}$ provides the best scores for PM_{10}. The averaged root-mean-square error (RMSE) of PM_{10} is 11.63 $\mu g\, m^{-3}$ with DA ($\mathrm{PM}_{2.5}$ and $\mathrm{PM}_{2.5-10}$), compared to 13.69 $\mu g\, m^{-3}$ with DA (PM_{10}) and 17.74 $\mu g\, m^{-3}$ without DA on 1 July 2009. The averaged RMSE of PM_{10} is 4.73 $\mu g\, m^{-3}$ with DA ($\mathrm{PM}_{2.5}$ and $\mathrm{PM}_{2.5-10}$), against 6.08 $\mu g\, m^{-3}$ with DA (PM_{10}) and 6.67 $\mu g\, m^{-3}$ without DA on 26 July 2009.

6.2.3. Assimilation of lidar signals: application to aerosol forecasting

Participants: Yiguo Wang [CEREA], Karine Sartelet [CEREA], Marc Bocquet, Patrick Chazette [LSCE].

This study represents a new application of assimilating lidar signals to aerosol forecasting. It aims at investigating the impact of a ground-based lidar network on the analysis and short-term forecasts of aerosols through a case study in the Mediterranean basin. To do so, we employ a Data Assimilation (DA) algorithm based on the optimal interpolation method developed in the Polair3D chemistry transport model (CTM) of the Polyphemus air quality modeling platform. We assimilate hourly averaged normalized range-corrected lidar signals retrieved from a 72 h period of intensive and continuous measurements performed in July 2012 by ground-based lidar systems of the European Aerosol Research Lidar Network (EARLINET). Particles with an aerodynamic diameter lower than 2.5 μm ($\mathrm{PM}_{2.5}$) and those with an aerodynamic diameter higher than 2.5 μm but lower than 10 ($\mathrm{PM}_{10-2.5}$) are analyzed separately using the lidar observations at each DA step. First, we study the spatial and temporal influences of the assimilation of lidar signals on aerosol forecasting. We conduct sensitivity studies on algorithmic parameters, e.g. the horizontal correlation length (L_h) used in the background error covariance matrix (50 km, 100 km or 200 km), the altitudes at which DA is performed (0.75–3.5 km, 1.0–3.5 km or 1.5–3.5 km) and the assimilation period length (12 h or 24 h). We find that DA with $L_\mathrm{h} = 100$ km and assimilation from 1.0 to 3.5 km during a 12 h assimilation period length leads to the best scores for PM_{10} and $\mathrm{PM}_{2.5}$ during the forecast period with reference to available measurements from surface networks. Secondly, the aerosol simulation results without and with lidar DA using the optimal parameters (L_h = 100 km, an assimilation altitude range from 1.0 to 3.5 km and a 12 h DA period) are evaluated using the level 2.0 (cloud-screened and quality-assured) aerosol optical depth data from AERONET, and mass concentration measurements (PM_{10} or $\mathrm{PM}_{2.5}$) from the French air quality (BDQA) network and the EMEP-Spain/Portugal network. The results show that the simulation with DA leads to better scores than the one without DA for $\mathrm{PM}_{2.5}$, PM_{10} and aerosol optical depth. Additionally, the comparison of model results to evaluation data indicates that the temporal impact of assimilating lidar signals is longer than 36 h after the assimilation period.

Fig. 2 shows the performance of assimilating real lidar data over the Mediterranean sea with a view to forecast particulate matter over France.

6.2.4. Local ensemble transform Kalman filter for adaptive optics on extremely large telescopes

Participants: Morgan Gray [LAM, France], Cyril Petit [ONERA, France], Sergei Rodionov [LAM, France], Marc Bocquet, Laurent Bertino [NERSC, Norway], Marc Ferrari [LAM, France], Thierry Fusco [LAM and ONERA, France].

We proposed a new algorithm for an adaptive optics system control law, based on the Linear Quadratic Gaussian approach and a Kalman Filter adaptation with localizations. It allows to handle non-stationary behaviors, to obtain performance close to the optimality defined with the residual phase variance minimization criterion, and to reduce the computational burden with an intrinsically parallel implementation on the Extremely Large Telescopes.

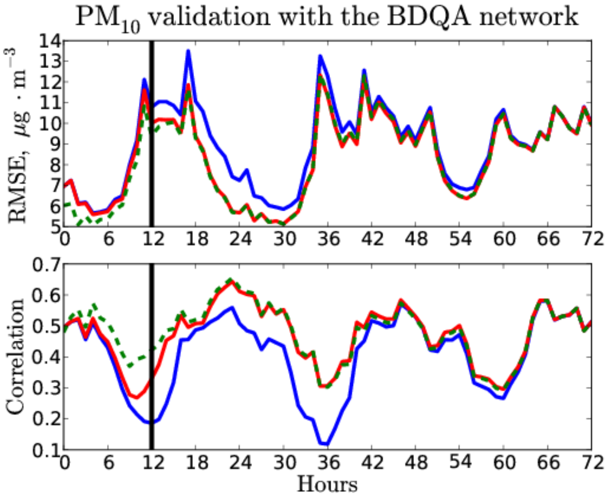


Figure 2. Validation of forecasts of particulate matter PM_{10} using ground stations over France when lidar data have been assimilated over the Mediterranean sea. These forecasts (red line: 12-hour assimilation period and dashed green line: 24-hour assimilation period) are compared to a free run (blue line).

6.3. Inverse modeling

Research on inverse modeling techniques is a major component of Clime, with a focus, in 2014, on hyperparameter estimation when the statistics are non-Gaussian.

6.3.1. Estimation of the caesium-137 source term from the Fukushima Daiichi plant

Participants: Victor Winiarek, Marc Bocquet, Nora Duhanyan [CEREA], Yelva Roustan [CEREA], Olivier Saunier [IRSN], Anne Mathieu [IRSN].

To estimate the amount of radionuclides and the temporal profile of the source term released in the atmosphere during the accident of the Fukushima Daiichi nuclear power plant in March 2011, inverse modeling techniques have been used and have proven their ability in this context. In a previous study, the lower bounds of the caesium-137 and iodine-131 source terms were estimated with such techniques, using activity concentration observations. The importance of an objective assessment of prior errors (the observation errors and the background errors) was emphasized for a reliable inversion. In such critical context where the meteorological conditions can make the source term partly unobservable and where only a few observations are available, such prior estimation techniques are mandatory, the retrieved source term being very sensitive to this estimation.

We propose to extend the use of these techniques to the estimation of prior errors when assimilating observations from several data sets. The aim is to compute an estimate of the caesium-137 source term jointly using all available data about this radionuclide, such as activity concentrations in the air, but also daily fallout measurements and total cumulated fallout measurements. It is crucial to properly and simultaneously estimate the background errors and the prior errors relative to each data set. A proper estimation of prior errors is also a necessary condition to reliably estimate the a posteriori uncertainty of the estimated source term. Using such techniques, we retrieve a total released quantity of caesium-137 in the interval $11.6 - 19.3$ PBq with an estimated standard deviation range of $15 - 20\%$ depending on the method and the data sets. The “blind” time intervals of the source term have also been strongly mitigated compared to the first estimations with only activity concentration data.

6.4. Image assimilation

Sequences of images, such as satellite acquisitions, display structures evolving in time. This information is recognized of major interest by forecasters (meteorologists, oceanographers, etc.) in order to improve the information provided by numerical models. However, the satellite images are mostly assimilated in geophysical models on a point-wise basis, discarding the space-time coherence visualized by the evolution of structures such as clouds. Assimilating in an optimal way image data is of major interest and this issue should be considered in two ways:

- from the model's viewpoint, the location of structures on the observations is used to control the state vector.
- from the image's viewpoint, a model of the dynamics and structures is built from the observations.

6.4.1. Model error and motion estimation

Participants: Dominique Béréziat [UPMC], Isabelle Herlin.

Data assimilation technics are used to retrieve motion from image sequences. These methods require a model of the underlying dynamics, displayed by the evolution of image data. In order to quantify the approximation linked to the chosen dynamic model, an error term is included in the evolution equation of motion and a weak formulation of 4D-Var data assimilation is designed. The cost function to be minimized depends simultaneously on the initial motion field, at the beginning of the studied temporal window, and on the error value at each time step. The result allows to assess the model error and analyze its impact on motion estimation. The approach is used to estimate geophysical forces (gravity, Coriolis, diffusion) from images in order to better assess the surface dynamics [20] and forecast the displacement of structures like oilspill.

6.4.2. Tracking of structures from an image sequence

Participants: Yann Lepoittevin, Isabelle Herlin, Dominique Béréziat [UPMC].

The research concerns an approach to estimate velocity on an image sequence and simultaneously segment and track a given structure. It relies on the underlying dynamics' equations of the studied physical system. A data assimilation method is designed to solve evolution equations of image brightness, those of motion's dynamics, and those of the distance map modeling the tracked structures. The method is applied on meteorological satellite data, in order to track tropical clouds on image sequences and estimate their motion, as seen on Fig. 3
.

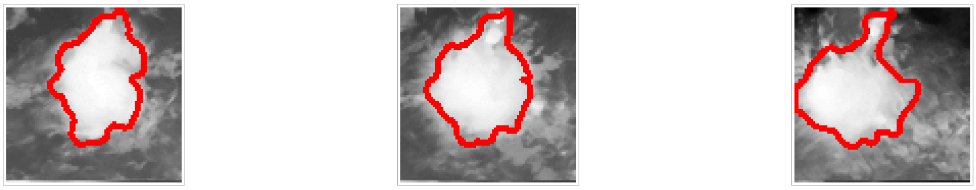


Figure 3. Tracking a tropical cloud. Frames 3, 9, 18 of the sequence.

Quantification is obtained on synthetic experiments by comparing trajectories of characteristic points. The respective position of these points on the last image of the sequence for different methods may be compared to that obtained with ground truth as seen on Fig. 4 .

Data assimilation is performed either with a 4D-Var variational approach or with a Kalman ensemble method [22]. In the last case, the initial ensemble is obtained from a set of optical flow methods of the literature with various parameters values.

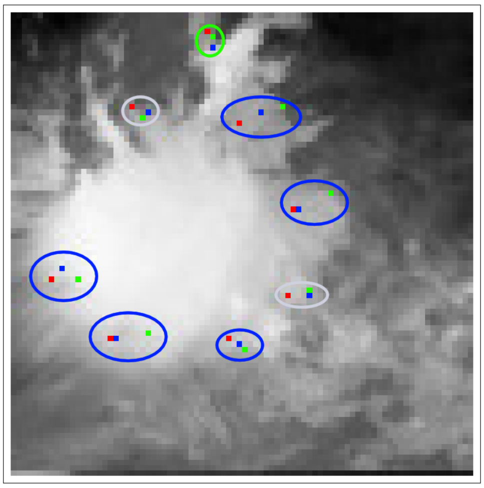


Figure 4. Red point: ground truth. Blue point: our method. Green point: Sun's optical flow. Blue ellipse: our method is the best. Green ellipse: Sun's result is the best. Grey ellipse : results are equivalent.

6.4.3. Motion estimation from images with a waveforms reduced model

Participants: Etienne Huot, Isabelle Herlin, Giuseppe Papari [CFLIR, Belgium].

Dimension reduction is applied to a model of image evolution, composed of transport of velocity and image brightness. Waveform bases are obtained on the image domain for subspaces of images and motion fields, as eigenvectors of previously defined quadratic functions. Image assimilation with the reduced model allows to estimate velocity fields satisfying the space-time properties chosen defined by the user for designing the quadratic function. This approach allows complex geographical domains and suppresses the difficulty of boundary conditions on such domains: these boundary conditions are automatically applied on the bases elements. Motion estimation is then obtained with a reduced model whose state vector is composed of a few components for motion and images. This has to be compared with the initial motion estimation problem that involves a state vector that has a size proportional to the image domain. Current research concern the definition of new quadratic functions from image properties.

6.4.4. Applying POD on a model output dabase for defining a reduced motion model

Participants: Etienne Huot, Isabelle Herlin.

Dimension reduction may also be studied by determining a small size reduced basis obtained by Proper Orthogonal Decomposition (POD) of a motion fields database. This database is constructed for characterizing accurately the surface circulation of the studied area, so that linear combinations of the basis elements obtained by POD accurately describe the motion function observed on satellite image sequences. The database includes the geostrophic motion fields obtained from Sea Level Anomaly reanalysis maps that are available from the MyOcean European project website (http://www.myocean.eu/). Fig. 5 displays such SLA maps and the associated motion fields.

Image assimilation with the POD reduced model allows estimating motion as displayed on Fig. 6 .

6.4.5. Rain nowcasting from radar image acquisitions

Participants: Yann Lepoittevin, Isabelle Herlin.

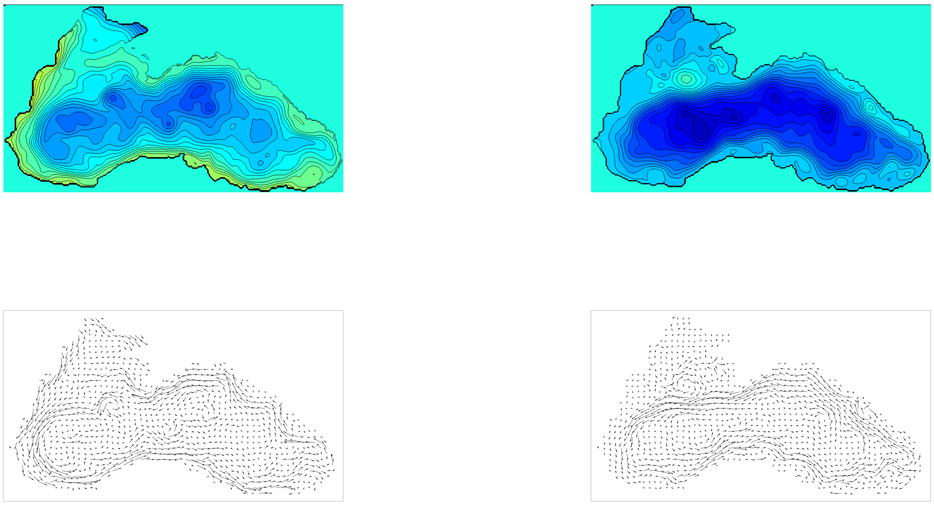


Figure 5. Top: reanalysis of SLA. Bottom: geostrophic motion.

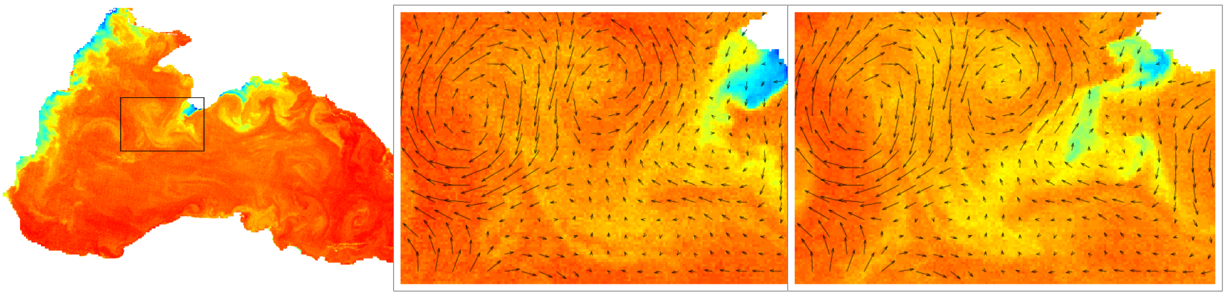


Figure 6. Zoom on a region of interest and motion estimation superposed on two consecutive images.

This research concerns the design of an operational method for rainfall nowcasting that aims at prevention of flash floods. The nowcasting method is based on two main components:

- a data assimilation method, based on radar images, estimates the state of the atmosphere: this is the estimation phase.
- a forecast method uses this estimation to extrapolate the state of the atmosphere in the future: this is the forecast phase.

Results were analyzed by Numtech (partner of a joint I-lab) on space-time neighborhood in order to prevent consequences of flash floods on previously defined zone.

Current research concerns the use of object components in the state vector in order to get an improved motion estimation and a better localization of endangered regions.

6.5. Uncertainty quantification and risk assessment

The uncertainty quantification of environmental models raises a number of problems due to:

- the dimension of the inputs, which can easily be 10^5-10^8 at every time step;
- the dimension of the state vector, which is usually 10^5-10^7;
- the high computational cost required when integrating the model in time.

While uncertainty quantification is a very active field in general, its implementation and development for geosciences requires specific approaches that are investigated by Clime. The project-team tries to determine the best strategies for the generation of ensembles of simulations. In particular, this requires addressing the generation of large multimodel ensembles and the issue of dimension reduction and cost reduction. The dimension reduction consists in projecting the inputs and the state vector to low-dimensional subspaces. The cost reduction is carried out by emulation, i.e., the replacement of costly components with fast surrogates.

6.5.1. Application of sequential aggregation to meteorology

Participants: Jean Thorey, Paul Baudin, Vivien Mallet, Stéphanie Dubost [EDF R&D], Christophe Chaussin [EDF R&D], Laurent Dubus [EDF R&D], Luc Musson-Genon [CEREA, EDF R&D], Laurent Descamps [Météo France], Philippe Blanc [Armines], Gilles Stoltz [CNRS].

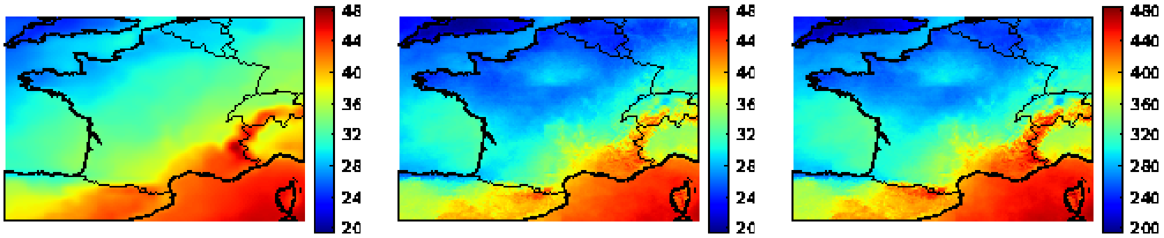
Nowadays, it is standard procedure to generate an ensemble of simulations for a meteorological forecast. Usually, meteorological centers produce a single forecast, out of the ensemble forecasts, computing the ensemble mean (where every model receives an equal weight). It is however possible to apply aggregation methods. When new observations are available, the meteorological centers also compute analyses. Therefore, we can apply the ensemble forecast of analyses. Ensembles of forecasts for mean sea level pressure, from the THORPEX Interactive Grand Global Ensemble, were aggregated with a forecast error decrease by 20% compared to the ensemble mean.

We studied the aggregation of ensembles of solar radiations in the context of photovoltaic production. The observations are based on MeteoSat Second Generation (MSG) and provided by the HelioClim-3 database as gridded fields. The ensembles of forecasts are from the THORPEX Interactive Grand Global Ensemble. The aggregated forecasts show a 20% error decrease compared to the individual forecasts. They are also able to retrieve finer spatial patterns than the ones found in the individual forecasts (see Figure 7).

6.5.2. Sequential aggregation with uncertainty estimation

Participants: Vivien Mallet, Jean Thorey, Paul Baudin, Gilles Stoltz [CNRS].

An important issue is the estimation of the uncertainties associated with the aggregated forecasts. We devised a new approach to predict a probability density function or cumulative distribution function instead of a single aggregated forecast. In practice, the aggregation procedure aims at forecasting the cumulative distribution function of the observations which is simply a Heaviside function centered at the observed value. Our forecast is the weighted empirical cumulative distribution function based on the ensemble of forecasts. The method guarantees that, in the long run, the forecast cumulative distribution function has a continuous ranked probability score at least as good as the best weighted empirical cumulative function with weights constant in time.



(a) (b) (c)

Figure 7. Yearly average of the map of downward shortwave solar radiation in Wm^{-2}*, for an ensemble mean (a), for our aggregated forecasts (b) and observed (c).*

6.5.3. Sensitivity analysis in the dispersion of radionuclides

Participants: Sylvain Girard, Vivien Mallet, Irène Korsakissok [IRSN].

We carried out a sensitivity analysis of the dispersion of radionuclides during Fukushima disaster. We considered the dispersion at regional scale, with the Eulerian transport model Polair3D from Polyphemus. The sensitivities to most input parameters were computed using the Morris method (with 8 levels and 100 trajectories). The influences of 19 scalar parameters were quantified. The scalar parameters were additive terms or multiplicative factors applied to 1D, 2D or 3D fields such as emission rates, precipitations, cloud height, wind velocity. The sensitivity analysis was carried out with the Morris method and by computing Sobol' indices. Both approaches were found to be consistent. Computing the Sobol' indices required the use of Gaussian process emulation, which proved to be successful at least on targets averaged in time and space.

It was shown that, depending on the output quantities of interest (various aggregated atmospheric and ground dose rates), the sensitivity to the inputs may greatly vary in time and space (see Figure 8). Very few parameters show low sensitivity in any case. The vertical diffusion coefficient, the scavenging factors, the winds and precipitation intensity were found to be the most influential inputs. Most input variables related to the source term (emission rates, emission dates) also had a strong influence.

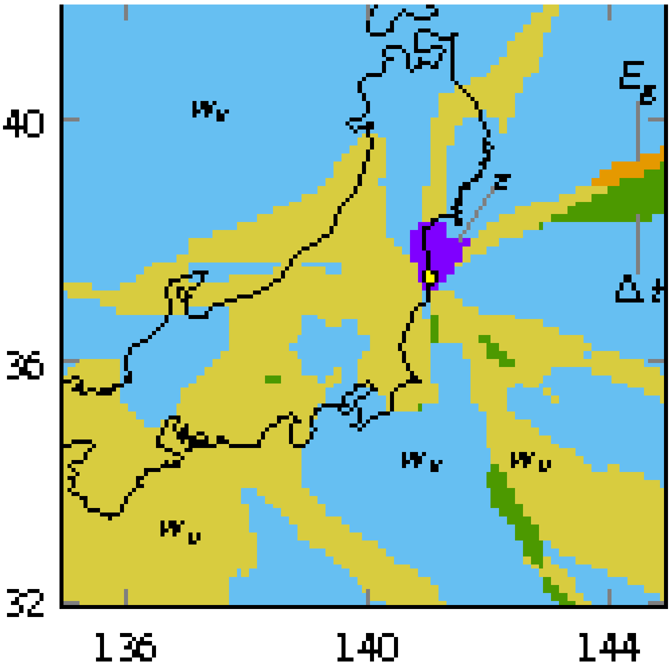


Figure 8. Variables that influence the most the atmospheric radioactivity after Fukushima disaster. z is the emissions altitude; Δt is the time shift on emissions; E_g stands for the emissions of noble gas; w_u and w_v are for zonal and meridional winds, respectively.

COFFEE Project-Team (section vide)

FLUMINANCE Project-Team

6. New Results

6.1. Highlights of the Year

6.1.1. Stochastic fluid flow dynamics under uncertainty

We have proposed the basis of a formalism allowing to built large scale stochastic representation of fluid flows dynamics [17]. This formalism relies on a location uncertainty principle which separates the flow in terms of a resolved large scale component and a highly oscillating random component. The dynamics is built in a similar way as in the deterministic case through a stochastic representation of the Reynolds transport theorem. This principle paves a new way for the construction of subgrid models from the uncertainties we have on the flow. The associated subgrid tensor provides a clear interaction between small scale data and large scale resolved quantities. This characteristic opens new directions for the devising of methods for the nulmerical simulation of large scale components of the flow. It allows also deriving large-scale models that takes into account explicitly the inherent errors to a particular geophysical dynamics representation.

6.2. Fluid motion estimation

6.2.1. Stochastic uncertainty models for motion estimation

Participants: Etienne Mémin, Manuel Saunier, Abed Malti.

In this study we have proposed a stochastic formulation of the brightness consistency used principally in motion estimation problems. In this formalization the image luminance is modeled as a continuous function transported by a flow known only up to some uncertainties. Stochastic calculus then enables to built conservation principles which take into account the motion uncertainties. These uncertainties defined either from isotropic or anisotropic models can be estimated jointly to the motion estimates. Such a formulation besides providing estimates of the velocity field and of its associated uncertainties allows us to naturally define a linear multiresolution scale-space framework. The corresponding estimator, implemented within a local least squares approach, has shown to improve significantly the results of the corresponding deterministic estimator (Lucas and Kanade estimator). This fast local motion estimator provides results that are of the same order of accuracy than state-of-the-art dense fluid flow motion estimator for particle images. The uncertainties estimated supply a useful piece of information in the context of data assimilation. This ability has been exploited to define multiscale incremental data assimilation filtering schemes. The development of an efficient GPU based version of this estimator recently started through the Inria ADT project FLUMILAB

6.2.2. 3D flows reconstruction from image data

Participants: Ioana Barbu, Kai Berger, Cédric Herzet, Etienne Mémin.

Our work focuses on the design of new tools for the estimation of 3D turbulent flow motion in the experimental setup of Tomo-PIV. This task includes both the study of physically-sound models on the observations and the fluid motion, and the design of low-complexity and accurate estimation algorithms. On the one hand, we investigate state-of-the-art methodologies such as ,“sparse representations" for the characterization of the observation and fluid motion models. On the other hand, we place the estimation problem into a probabilistic Bayesian framework and use state-of- the-art inference tools to effectively exploit the strong time-dependence on the fluid motion.

Last year, we focused on the design of new methodologies to jointly estimate the volume of particles and the velocity field from the received image data. Our approach was based on the minimization (with respect to both the position of the particles and the velocity field) of a cost function penalizing both the discrepancies with respect to a conservation equation and some prior estimates of particle positions.

This year, we revisited the problem of volume reconstruction through the prism of some modern optimization techniques. More specifically, we focussed our attention on the family of proximal and splitting methods and showed that the standard techniques commonly adopted in the TomoPIV literature can be seen as particular cases of such methodologies. Recasting standard methodologies in a more general framework allowed us to propose extensions of the latter: i) we showed that the parcimony characterizing the sought volume can be accounted for without increasing the complexity of the algorithms (e.g., by including simple thresholding operations); ii) we emphasized that the speed of convergence of the standard reconstruction algorithms can be improved by using Nesterov's acceleration schemes; iii) we also proposed a totally novel way of reconstructing the volume by using the so-called “alternating direction of multipliers method" (ADMM) . The journal publications relative to the contributions developped this year are currently in construction.

6.2.3. *Sparse-representation algorithms*

Participant: Cédric Herzet.

The paradigm of sparse representations is a rather new concept which turns out to be central in many domains of signal processing. In particular, in the field of fluid motion estimation, sparse representation appears to be potentially useful at several levels: i) it provides a relevant model for the characterization of the velocity field in some scenarios; ii) it plays a crucial role in the recovery of volumes of particles in the 3D Tomo-PIV problem.

Unfortunately, the standard sparse representation problem is known to be NP hard. Therefore, heuristic procedures have to be devised to access to the solution of this problem. Among the popular methods available in the literature, one can mention orthogonal matching pursuit (OMP), orthogonal least squares (OLS) and the family of procedures based on the minimization of ℓ_p norms. In order to assess and improve the performance of these algorithms, theoretical works have been undertaken in order to understand under which conditions these procedures can succeed in recovering the "true" sparse vector.

Last, we contributed to this research axis by deriving conditions of success for the algorithms mentioned above when some partial information is available about the position of the nonzero coefficients in the sparse vector. This paradigm is of interest in the Tomographic-PIV volume reconstruction problem: one can indeed expect volumes of particles at two successive instants to be pretty similar; any estimate of the position of the particles at one given instant can therefore serve as a prior estimate about their position at the next instant. Another information of interest which can help the algorithms in their reconstruction process is the decay of the amplitude of the nonzero coeffcient in the sparse vector. In a TomoPIV context, this decay corresponds to the fact that not all the particles in fluid diffuse the same quantity of light (notably beacuse of illumination or radius variation). This year, we thus pursue our effort in the understanding of the success of some reconstruction algorithms when the sparse vectors obey some decay. In particular, we showed that the standard coherence-based guarantees for OMP/OLS can be relaxed by an amount which depends on the decay of the nonzero coeffcients.

Another axis of research we have dealt with is the extension of sparse methodologies to the context of nonlinear models. This type of situtation is indeed frequently encountered in fluid mechanics or geophysics where the initial/boundary conditions of a system are known to be sparse in some basis and the collected observations obey a nonlinear dynamical model (e.g., the Navier-Stokes equations). In our work, we showed that many sparse representation algorithms, designed in the linear paradigm, can be nicely extended to the nonlinear setup provided that the gradient of the functional can be evaluated efficiently. In order to do so, we suggested a methodology, well-known in the commmunity of optimal control, but surprinsingly quite uncommon in many fields of signal processing.

Our work have led to the publication of contributions in the IEEE International Conference on Speech, Acoustic and Signal Processing (ICASSP) [23] and international - Traveling Workshop on Interactions between Sparse models and Technology (iTwist) [22],[24]

6.3. Tracking, Data assimilation and model-data coupling

6.3.1. Stochastic filtering technique for the tracking of closed curves

Participant: Etienne Mémin.

We have studied a stochastic filtering technique for the tracking of closed curves along an image sequence. In that goal, we designed a continuous-time stochastic dynamics that allows us to infer inter-frame deformations. The curve is defined by an implicit level-set representation and the stochastic dynamics is expressed properly on the level-set function. It takes the form of a stochastic partial differential equation with a Brownian motion of low dimension. The evolution model we proposed combines local photometric information, deformations induced by the curve displacement and an uncertainty modeling of the dynamics. Specific choices of noise models and drift terms lead to an evolution law based on mean curvature as in classic level set methods, while other choices yield new evolution laws. The approach we propose is implemented through a particle filter, which includes color measurements characterizing the target and the background photometric probability densities respectively. The merit of this parameter free filter is demonstrated on various satellite image sequences depicting the evolution of complex geophysical flows. This work has been recently published in the Journal of Mathematical Imaging and Vision [13]. Let us note the method provides an empirical dynamical model learned recursively from a data flow. Its short time forecasting skills have been used in the context of weather-watch radar images within a fruitful collaboration with MeteoFrance.

6.3.2. Sequential smoothing for fluid motion

Participants: Anne Cuzol, Etienne Mémin.

In parallel to the construction of stochastic filtering techniques for fluid motions, we have proposed a new sequential smoothing method within a Monte-Carlo framework. This smoothing aims at reducing the temporal discontinuities induced by the sequential assimilation of discrete time data into continuous time dynamical models. The time step between observations can indeed be long in environmental applications for instance, and much longer than the time step used to discretize the model equations. While the filtering aims at estimating the state of the system at observations times in an optimal way, the objective of the smoothing is to improve the estimation of the hidden state between observation times. The method is based on a Monte-Carlo approximation of the filtering and smoothing distributions, and relies on a simulation technique of conditional diffusions. The proposed smoother can be applied to general non linear and multidimensional models. It has been applied to a turbulent flow in a high-dimensional context, in order to smooth the filtering results obtained from a particle filter with a proposal density built from an Ensemble Kalman procedure. This conditional simulation framework can also be used for filtering problem with low measurement noise. This has been explored through a collaboration with Jean-Louis Marchand (ENS Bretagne) in the context of vorticity tracking from image data.

6.3.3. Stochastic fluid flow dynamics under uncertainty

Participants: Etienne Mémin, Valentin Resseguier.

In this research axis we aim at devising Eulerian expressions for the description of fluid flow evolution laws under uncertainties. Such an uncertainty is modeled through the introduction of a random term that allows taking into account large-scale approximations or truncation effects performed within the dynamics analytical constitution steps. This includes for instance the modeling of unresolved scales interaction in large eddies simulation (LES) or in Reynolds average numerical simulation (RANS), but also uncertainties attached to non-uniform grid discretization. This model is mainly based on a stochastic version of the Reynolds transport theorem. Within this framework various simple expressions of the drift component can be exhibited for different models of the random field carrying the uncertainties we have on the flow. We aim at using such a formalization within image-based data assimilation framework and to derive appropriate stochastic versions of geophysical flow dynamical modeling. This formalization has been published in the journal Geophysical and Astrophysical Fluid Dynamics [17]. Numerical simulation on divergence free wavelets basis of 3D viscous Taylor-Green vortex and Crow instability have been performed within a collaboration with Souleymane Kadri-Harouna. First promising results have been obtained. Besides, we explore in the context of Valentin Resseguier's PhD the extension of such framework to oceanic models and to satellite image data assimilation. This PhD thesis takes place within a fruitful collaboration with Bertrand Chapron (CERSAT/IFREMER).

6.3.4. Free surface flows reconstruction and tracking

Participants: Dominique Heitz, Etienne Mémin.

We investigated the combined use of a Kinect depth sensor and of a stochastic data assimilation method to recover free-surface flows. More generally, we proposed a particle filter method to reconstruct the complete state of free-surface flows from a sequence of depth images only. The data assimilation scheme introduced accounts for model and observations errors. We evaluated the developed approach on two numerical test cases: a collapse of a water column as a toy-example and a flow in an suddenly expanding flume as a more realistic flow. The robustness of the method to simulated depth data quality and also to initial conditions was considered. We illustrated the interest of using two observations instead of one observation into the correction step. Then, the performance of the Kinect sensor to capture temporal sequences of depth observations was investigated. Finally, the efficiency of the algorithm was qualified for a wave in a real rectangular flat bottom tank. It was shown that for basic initial conditions, the particle filter rapidly and remarkably reconstructed velocity and height of the free surface flow based on noisy measures of the elevation alone. These results have been recently submitted to a special issue of Fluid Dynamics Research.

6.3.5. Variationnal ensemble methods for data assimilation

Participants: Dominique Heitz, Etienne Mémin, Cordelia Robinson, Yin Yang.

In this work, we aim at studying an ensemble based optimal control strategy for data assimilation. Such a formulation nicely combines the ingredients of ensemble Kalman filters and variational assimilation. In the same way as standard variational assimilation, it is formulated as the minimization of an objective function. However, similarly to ensemble filters, it introduces in its objective function an empirical ensemble-based background-error covariance and works in an off-line smoothing mode rather than sequentially like filtering approaches in a sequential filter. These techniques have the great advantage to avoid the constitution of tangent linear and adjoint models, which are necessary for standard incremental variational techniques. As the background error covariance matrix plays a key role in the variational process, our study particularly focuses on the generation of the analysis ensemble state with localization techniques. The proposed method was assessed with a Shallow Water model combined with synthetic data and original incomplete experimental depth sensor observations. Results submitted to Computers & Fluids showed that the modified ensemble technique was better in quality and reduced the computational cost.

6.3.6. Optimal control techniques for the coupling of large scale dynamical systems and image data

Participants: Dominique Heitz, Etienne Mémin, Cordelia Robinson.

This work aims at investigating the use of optimal control techniques for the coupling of Large Eddies Simulation (LES) techniques and 2D image data. The objective is to reconstruct a 3D flow from a set of simultaneous time resolved 2D image sequences visualizing the flow on a set of 2D plans enlightened with laser sheets. This approach will be experimented on shear layer flows and on wake flows generated on the wind tunnel of Irstea Rennes. Within this study we wish also to explore techniques to enrich large-scale dynamical models by the introduction of uncertainty terms or through the definition of subgrid models from the image data. This research theme is related to the issue of turbulence characterization from image sequences. Instead of predefined turbulence models, we aim here at tuning from the data the value of coefficients involved in traditional LES subgrid models or in longer-term goal to learn empirical subgrid models directly from image data. An accurate modeling of this term is essential for Large Eddies Simulation as it models all the non resolved motion scales and their interactions with the large scales.

We have pursued the first investigations on a 4DVar assimilation technique, integrating PIV data and Direct Numerical Simulation (DNS), to reconstruct two-dimensional turbulent flows. The problem we are dealing with consists in recovering a flow obeying Navier-Stokes equations, given some noisy and possibly incomplete PIV measurements of the flow. By modifying the initial and inflow conditions of the system, the proposed method reconstructs the flow on the basis of a DNS model and noisy measurements. The technique has been evaluated in the wake of a circular cylinder. It denoises the measurements and increases the spatiotemporal resolution of PIV time series. These results have been recently published in the Journal of Computational Physics [6]. Along the same line of studies the 3D case is ongoing. The goal consists here to reconstruct a 3D flow from a set of simultaneous time resolved 2D images of planar sections of the 3D volume. This work is mainly conducted within the PhD of Cordelia Robinson. The development of the variational assimilation code has been initiated within a collaboration with A. Gronskis, S. Laizé (lecturer, Imperial College, UK) and Eric Lamballais (institut P' Poitiers). A High Reynolds number simulation of the wake behind a cylinder has been recently performed within this collaboration. The 4DVar assimilation technique based on the numerical code Incompact3D is now implemented. We are currently trying to reconstruct a 3D turbulent flow from dual plane velocity observations. First assessments have been carried out with DNS based synthetic data. Further evaluation will be done with real measurements based on dual stereo PIV experiments.

6.3.7. *Ensemble variational data assimilation of large scale fluid flow dynamics with uncertainty*

Participants: Etienne Mémin, Yin Yang.

In this work we explore the assimilation of a large scale representation of the flow dynamics with image data of finer resolution. The velocity field at large scales is described as a regular smooth component whereas the complement component is a highly oscillating random velocity field defined on the image grid but living at all the scales. Following this route we have started to assess the performance of a variational assimilation technique with direct image data observation. Preliminary encouraging results have been obtained for simulation under uncertainty of 1D and 2D shallow water models.

6.3.8. *Reduced-order models for flows representation from image data*

Participants: Cédric Herzet, Etienne Mémin.

One of the possibilities to neglect the influence of some degrees of freedom over the main characteristics of a flow consists in representing it as a sum of K orthonormal spatial basis functions weighted with temporal coefficients. To determine the basis function of this expansion, one of the usual approaches relies on the Karhunen-Loeve decomposition (refered to as proper orthogonal decomposition – POD – in the fluid mechanics domain). In practice, the spatial basis functions, also called modes, are the eigenvectors of an empirical auto-correlation matrix which is built from “snapshots" of the considered physical process.

In this axis of work we focus on the case where one does not have a direct access to snapshots of the considered physical process. Instead, the POD has to be built from the partial and noisy observations of the physical phenomenon of interest. Instances of such scenarios include situations where real instantaneous vector-field snapshots are estimated from a sequence of images. We have been working on several approaches dealing with such a new paradigm. A first approach consists in extending standard penalized motion-estimation algorithms to the case where the sought velocity field is constrained to span a low-dimensional subspace. In particular, we have considered scenarios where the standard optical flow constraint (OFC) is no longer satisfied and one has therefore to resort to a Discrete Finite Difference (DFD) model. The non-linearity of the latter leads to several practical issues that we have addressed this year.

6.4. Analysis and modeling of turbulent flows

6.4.1. *Hot-wire anemometry at low velocities*

Participant: Dominique Heitz.

A new dynamical calibration technique has been developed for hot-wire probes. The technique permits, in a short time range, the combined calibration of velocity, temperature and direction calibration of single and multiple hot-wire probes. The calibration and measurements uncertainties were modeled, simulated and controlled, in order to reduce their estimated values. Based on a market study the french patent application has been extended this year to a Patent Cooperation Treaty (PCT) application.

6.4.2. *Numerical and experimental image and flow database*

Participant: Dominique Heitz.

The goal was to design a database for the evaluation of the different techniques developed in the Fluminance group. The first challenge was to enlarge a database mainly based on two-dimensional flows, with three-dimensional turbulent flows. Synthetic image sequences based on homogeneous isotropic turbulence and on circular cylinder wake have been provided. These images have been completed with time resolved Particle Image Velocimetry measurements in wake and mixing layers flows. This database provides different realistic conditions to analyse the performance of the methods: time steps between images, level of noise, Reynolds number, large-scale images. The second challenge was to carried out orthogonal dual plane time resolved stereoscopic PIV measurements in turbulent flows. The diagnostic employed two orthogonal and synchronized stereoscopic PIV measurements to provide the three velocity components in planes perpendicular and parallel to the streamwise flow direction. These temporally resolved planar slices observations will be used in 4DVar assimilation technique, integrating Direct Numerical Simulation (DNS) and Large Eddies Simulation (LES), to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson. The third challenge was to carried out a time resolved tomoPIV experiments in a turbulent wake flow. These temporally resolved volumic observations will be used to assess the algorithms developped in the PhD of Ioana Barbu and in the postdoc of Kai Berger. Then this data will be used in 4DVar assimilation technique to reconstruct three-dimensional turbulent flows. This reconstruction will be conducted within the PhD of Cordelia Robinson.

6.5. Visual servoing approach for fluid flow control

6.5.1. Minimization of the kinetic energy density in the 2D plane Poiseuille flow

Participants: Christophe Collewet, Xuan Quy Dao.

This works concerns the PhD thesis of Xuan-Quy Dao. This year we have focused on a way to ensure a strict decreasing of the kinetic energy density. In that purpose, we have first proposed an approach to increase the controlled degrees of freedom. Indeed, the classical way to model this flow leads to only two degrees of freedom. With so few degrees of freedom it is obviously impossible to reach high desired performances as the strict minimization of the kinetic energy density. This way to proceed leads to a better minimization of the kinetic energy density. We have also proposed on approach based on a local decoupling of the controlled degree of freedom of the system so that an exponential decoupled decrease of each components of the state vector is locally obtained.

6.5.2. Control of systems described by partial differential equations

Participants: Tudor-Bogdan Airimitoaie, Christophe Collewet.

This work concerns principally the post-doctoral research of Tudor-Bogdan Airimiţoaie. It aims at controlling continuously evolving systems described by partial differential equations (PDEs). This is relevant in the context of the Fluminance team because fluid flows are infinite dimensional systems and can be rigorously described only through PDEs. In spite of this, practical approaches of flow control are based on low order numerical implementation relying on space and time discretization of the continuous system. This implies to setup strategies for model reduction that must be then in return properly understood with respect to the convergence of the control law. For finite dimensional implementations, one of the research directions pursued concerns the study on the benefit of increasing the controlled degrees of freedom (see the work of Xuan-Quy Dao). Another research direction, started recently, consists in improving control by using real-time estimation of a finite number of parameters related to the original infinite dimensional system. Indeed, this opens the possibility of improving performances by using more advanced robust linear parametric varying (LPV) control techniques existing in the literature. Two conference papers on these works have been submitted at the "7th AIAA Flow Control Conference".

KALIFFE Project-Team

6. New Results

6.1. Mixed semi-Lagrangian/finite difference methods for plasma simulations

We present an efficient algorithm for the long time behavior of plasma simulations. We will focus on 4D drift-kinetic model, where the plasma's motion occurs in the plane perpendicular to the magnetic field and can be governed by the 2D guiding-center model.

Hermite WENO reconstructions are applied for solving the Vlasov equation. Here we consider an arbitrary computational domain with an appropriate numerical method for the treatment of boundary conditions.

Then we apply this algorithm for plasma turbulence simulations. We first solve the 2D guiding-center model in a D-shape domain and investigate the numerical stability of the steady state. Then, the 4D drift-kinetic model is studied with a mixed method, *i.e.* the semi-Lagrangian method in linear phase and finite difference method during the nonlinear phase. Numerical results show that the mixed method is efficient and accurate in linear phase and it is much stable during the nonlinear phase. Moreover, in practice it has better conservation properties [9].

6.2. High order semi implicit schemes for PDEs

We consider a new formulation of implicit-explicit (IMEX) methods for the numerical discretization of time dependent partial differential equations. We construct several semi-implicit Runge-Kutta methods up to order three. This method is particularly well suited for problems where the stiff and non-stiff components cannot be well separated. We present different numerical simulations for reaction-diffusion, convection diffusion and nonlinear diffusion system of equations. Finally, we conclude by a stability analysis of the schemes for linear problems.

6.3. A Hierarchy of Hybrid Numerical Methods for Multi-Scale Kinetic Equations

We construct a hierarchy of hybrid numerical methods for multi-scale kinetic equations based on moment realizability matrices, a concept introduced by Levermore, Morokoff and Nadiga. Following such a criterion, one can consider hybrid scheme where the hydrodynamic part is given either by the compressible Euler or Navier-Stokes equations, or even with more general models, such as the Burnett or super-Burnett systems [8].

6.4. Derivation of high order absorbing boundary conditions for the Helmholtz equation in 2D

We present high order absorbing boundary conditions (ABC)for the Helmholtz equation in 2D, that can adapt to any regular shapedsurfaces. The new ABCs are derived by using the technique ofmicro-diagonalisation to approximate the Dirichlet-to-Neumann map.Numerical results on different shapes illustrate the behavior of thenew ABCs along with high-order finite elements [5].

LEMON Team

6. New Results

6.1. Highlights of the Year

Antoine ROUSSEAU and 5 co-authors released in 2014 the book *Brèves de Maths* [16]. This work (in french) selected more than 100 posts from the blog breves-de-maths.fr, in the framework of the international initiative "Mathematics of the Planet Earth". In this book (see cover 5), no complicated numbers, no weird equation, but short and clear sentences together with nice drawings to illustrate everyday life topics on our planet with the beauty of mathematics.

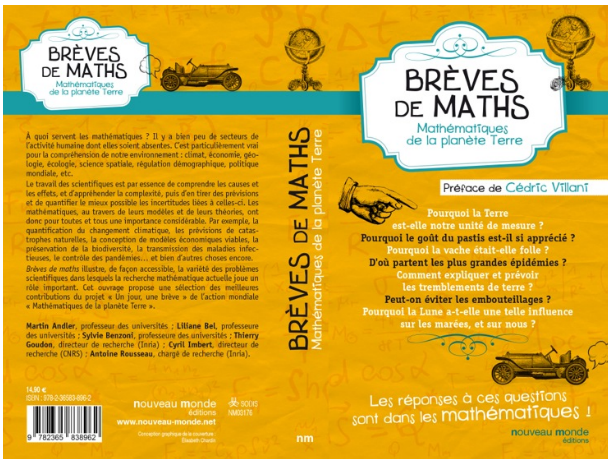


Figure 5. Brèves de Maths. Ed. Nouveau Monde, 2014

6.2. A Schwarz coupling method for dimensionally heterogeneous problem

Participant: Antoine Rousseau.

We study and analyze in [10] an efficient iterative coupling method for a dimensionally heterogeneous problem. We consider the case of 2-D Laplace equation with non symmetric boundary conditions with a corresponding 1-D Laplace equation. We first show how to obtain the 1-D model from the 2-D one by integration along one direction, by analogy with the link between shallow water equations and the Navier-Stokes system. Then, we focus on the design of an Schwarz-like iterative coupling method. We discuss the choice of boundary conditions at coupling interfaces. We prove the convergence of such algorithms and give some theoretical results related to the choice of the location of the coupling interface, and the control of the difference between a global 2-D reference solution and the 2-D coupled one. These theoretical results are illustrated numerically.

6.3. Bioremediation of water ressources

Participants: Antoine Rousseau, Alexis Pacholik.

Together with fellows from the MODEMIC team, we proposed a strategy for the bioremediation of water ressources such as lakes or lagoons. The originality is that the water treatment has to be done outside of the resource, in order not to dislocate its fragile ecological equilibrium.

The objective is to reach a targeted acceptable state for the resource within the minimal time. The patent [19] has been filed in connection with this work.

6.4. A well-balanced and positive preserving DG scheme for the SW equations

Participants: Arnaud Duran, Fabien Marche.

We consider in [5] the discontinuous Galerkin discretization of the nonlinear Shallow Water equations on unstructured triangulations. We propose an efficient combination of ingredients that leads to a simple high-order robust and well-balanced scheme, based on the alternative formulation of the equations known as the pre-balanced shallow water equations. We show that the preservation of the motionless steady states can be achieved, for an arbitrary order of polynomial expansion. Additionally, the preservation of the positivity of the water height is ensured using the recent method introduced in [59]. Some comparisons with a recent finite-volume MUSCL approach are also performed. The well-known tsunami test case shown in figures 6 and 7 has been computed here with high order DG scheme on unstructured triangulation.

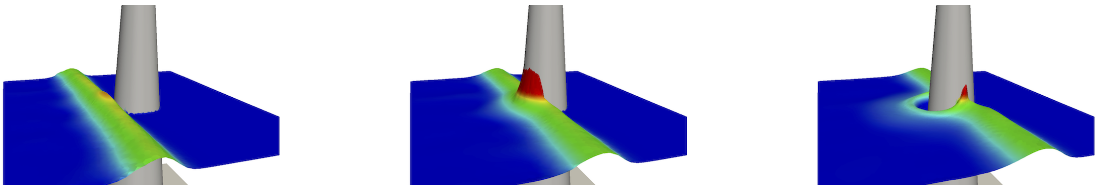


Figure 6. Tsunami wave over a conical island - Lateral view of the free surface at times t=5,6 and 7s.

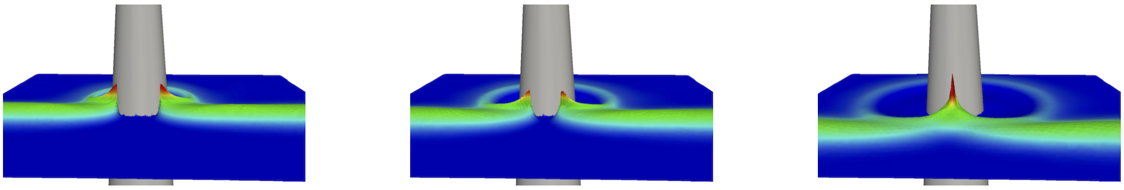


Figure 7. Tsunami wave over a conical island - Rear view of the free surface at times t=8,9 and 10s.

6.5. A well-balanced and positive preserving DG scheme for the GN equations

Participants: Arnaud Duran, Fabien Marche.

We introduce in [4] a discontinuous-Galerkin Finite-Element method to approximate the solutions of a new family of 1d Green-Naghdi models. These new models are shown to be more computationally efficient, while being asymptotically equivalent to the initial formulation with regard to the shallowness parameter. Using the free surface instead of the water height as a conservative variable, the models are recasted under a pre-balanced formulation and discretized using a nodal expansion basis. Independently from the polynomial degree in the approximation space, the preservation of the motionless steady-states is automatically ensured, and the water height positivity is enforced. A simple numerical procedure devoted to stabilize the computations in the vicinity of broken waves is also described. The validity of the resulting model is assessed through extensive numerical validations.

6.6. A new class of fully nonlinear and weakly dispersive Green-Naghdi models

Participant: Fabien Marche.

We introduce in [8] a new class of two-dimensional fully nonlinear and weakly dispersive Green-Naghdi equations over varying topography. These new Green-Naghdi systems share the same order of precision as the standard one but have a mathematical structure which makes them much more suitable for the numerical resolution, in particular in the demanding case of two dimensional surfaces. For these new models, we develop a high order, well balanced, and robust numerical code relying on a hybrid finite volume and finite difference splitting approach. The hyperbolic part of the equations is handled with a high-order finite volume scheme allowing for breaking waves and dry areas. The dispersive part is treated with a finite difference approach. Higher order accuracy in space and time is achieved through WENO reconstruction methods and through an SSP-RK time stepping. Particular effort is made to ensure positivity of the water depth.

6.7. Upscaling transfer properties in heterogeneous porous media

Participant: Vincent Guinot.

In [9] the passive solute transport was studied in a periodic, artificial porous medium. A Laplace analysis of the breakthrough curves indicates that the widely used, classical Advection-Dispersion (AD) model cannot reproduce the contaminant transport features accurately. Neither can fractional dynamics-based, anomalous dispersion models. The models failing to reproduce the features of contaminant transport is shown to be due to the Fick-like, gradient-based operator used to represent dispersion, that induces infinite signal propagation speed, even when fractional models are used. The Laplace analysis shows that advection processes are predominant at all time and space scales. The size of the Representative Elementary Volume is shown to be 20 to 30 periods.

MAGIQUE-3D Project-Team

6. New Results

6.1. Inverse Problems

6.1.1. Complex-frequency domain Full Waveform Inversion

Participants: Florian Faucher, Maarten V. de Hoop, Henri Calandra.

We study the seismic inverse problem for the (complex) frequency-domain elastic isotropic wave equation; in particular the recovery of the Lamé parameters and density. We employ a Full Waveform Inversion where the iterative minimization is based on a gradient descent. The elastic inverse problem shows a Lipschitz-type stability where the Fréchet derivative has a strictly positive 'lower bound'. This bound is connected to the stability constant and can be approximated using the Gauss-Newton Hessian. The successive stability estimates provide a control of the convergence and decide the parameters of inversion. We develop a multi-level approach based on a structured domain partitioning of the sub-surface. The coefficients (Lamé parameters and density) are assumed to be piecewise constant functions following the domain partitioning, which is naturally defined with the successive stability estimates to maintain the radius of convergence, while refinement provides resolution. It allows us to start with minimal prior information for the coefficients and the algorithm is perfectly suitable for complex frequency. We have carried out numerical experiments in two and three dimensions; those results have been presented during the following conferences in 2014: [48], [49].

6.1.2. Imaging of complex media with elastic wave equations

Participants: Jérôme Luquel, Hélène Barucq, Henri Calandra, Julien Diaz.

Even if RTM has enjoyed the tremendous progresses of scientific computing, its performances can still be improved, in particular when applied to strong heterogeneous media. In this case, images have been mainly obtained by using direct arrivals of acoustic waves and the transition to elastic waves including multiples is not obvious essentially because elastic waves equations are still more computationally consuming. The accuracy of numerical wave fields is obviously of great importance. We have thus chosen to consider high-order Discontinuous Galerkin Methods which are known to be well-adapted to provide accurate solutions based upon parallel computing. Now one of the main drawback of RTM is the need of storing a huge quantity of information which is redhibitory when using elastic waves. For that purpose, we apply the Griewank algorithm following Symes' ideas for the acoustic RTM. The idea is to find a compromise between the number of wave equations to solve and the number of numerical waves that we have to store. This is the so-called Optimal Checkpointing. By reducing the occupancy of the memory, RTM should be efficient even when using elastic waves. By this way, one may wonder if considering elastic waves including multiples in order to improve images of heterogeneous media is a valid option. It must involve a careful numerical analysis including the evaluation of the impact of the imaging condition. It is thus necessary to derive accurate imaging conditions, which could take advantage of all the information contained in the wavefield. For acoustic media, Claerbout proposed an imaging condition which is widely used and turns out to be sufficient to accurately reproduce interfaces. But Claerbout conditions do not take wave conversions into account and it is not clear wether conversions do or do not contain interesting information to get accurate images of heterogeneous media.

Since P-wave and S-wave interact with each other, it might be relevant to use an imaging condition including these interactions. In fact, this has been done successfully by J.Tromp and C. Morency for seismology applications based upon the inversion of the global Earth. Their approach is based upon the state adjoint and it involves sensitivity kernels which are defined from the propagated and the back-propagated fields. Now it has been shown that full wave form inversions using these sensitivity kernels may be polluted by numerical artefacts. One solution is to use a linear combination of the sensitivity kernels to delete artefacts. In this work, we propose then a new imaging condition which construction is inspired from with some approximations required to keep admissible computational costs. We illustrate the properties of the new imaging condition on industrial benchmarks like the Marmousi model. In particular, we compare the new imaging condition with other imaging conditions by using as criteria the quality of the image and the computational costs required by the RTM. This work was presented at the the WCCM XI - ECCM V - ECFD VI - Barcelona 2014 Conference and SIAM Conference on IMAGING SCIENCE (SIAM-IS14) Hong Kong Baptist University[67].

6.1.3. Helioseismology

Participants: Juliette Chabassier, Marc Duruflé, Thorsten Hohage.

We have begun to write a software interface that allows to solve an inverse problem using adjoint and regularization methods (iTReg software) while using Montjoie software for the direct problem that must be solved at each iteration of the inversion process.

6.2. Modeling

6.2.1. High-Order Time Schemes

6.2.1.1. Fourth order energy-preserving locally implicit discretization for linear wave equations

Participants: Juliette Chabassier, Sébastien Imperiale.

A family of fourth order coupled implicit-explicit schemes is presented as a special case of fourth order coupled implicit schemes for linear wave equations. The domain of interest is decomposed into several regions where different fourth order time discretization are used, chosen among a family of implicit or explicit fourth order schemes derived in [72]. The coupling is based on a Lagrangian formulation on the boundaries between the several non conforming meshes of the regions. A global discrete energy is shown to be preserved and leads to global fourth order consistency. Numerical results in 1d and 2d illustrate the good behavior of the schemes and their potential for the simulation of realistic highly heterogeneous media or strongly refined geometries, for which using everywhere an explicit scheme can be extremely penalizing. Accuracy up to fourth order reduces the numerical dispersion inherent to implicit methods used with a large time step, and makes this family of schemes attractive compared to second order accurate methods in time. This work has been presented at the Franco-Russian workshop on mathematical geophysics, Sep 2014, Novosibirsk, Russia [58], at the and is the object of a submitted publication to International Journal for Numerical Methods in Engineering.

6.2.1.2. A new modified equation approach for solving the wave equation

Participants: Hélène Barucq, Henri Calandra, Julien Diaz, Florent Ventimiglia.

In order to obtain high-order time-schemes, we are considering an alternative approach to the ADER schemes and to the modified equation technique described in section 3.2 . The two first steps of the construction of the schemes are similar to the previous schemes : we apply a Taylor expansion in time to the solution of the wave equation and we replace the high-order derivatives with respect to the time by high order space operators, using the wave equation. The difference is that we do not use auxiliary variables and we choose to discretize directly the high-order operators in space.

In the framework of the PhD thesis of Florent Ventimiglia, we have extended this new method involving p-harmonic operator to the first order formulation of the acoustic wave equation, which is the formulation discretized in the DIVA platform of TOTAL. In this case, the high order operators in space are not are not powers of the Laplace operator but powers of the gradient. Hence, we also had to adapt the space discretization, and we have extended the DG formulation with centered fluxes proposed in [77] to higher order operators. A numerical analysis of performance in 2D indicates that, for a given accuracy, this method requires less computational costs and less storage than the High-Order ADER Scheme. These results have been presented to the AIMS conference [54]. A paper has been published in ESAIM Proceedings [19].

6.2.2. Finite Element Methods for the time-harmonic wave equation.

6.2.2.1. Goal-Oriented Adaptivity using Unconventional Error Representations

Participants: Vincent Darrigrand, David Pardo, Ignacio Muga, Hélène Barucq.

In the scope of subsurface modelling via the resolution of inverse problems, the so-called goal-oriented adaptivity plays a fundamental role. Indeed, while classical adaptive algorithms were first designed to accurately approximate the energy norm of a problem [69], [70], one requires a good approximation of a specific quantity of interest. An energy norm driven self-adaptive strategy can still be used for that purpose, although it often becomes sub-optimal and unable to provide an accurate solution for the required quantity of interest in a reasonable amount of time.

During the late 90's, to overcome this issue, the so-called goal-oriented strategy appeared, see for instance [82], [81]. The goal-oriented approach consists in expressing the error in the quantity of interest as an integral over the entire computational domain involving the errors of the original and adjoint problems, and then minimise an upper bound of such error representation by performing local refinements.

Most authors, using the adjoint problem, represent the approximation error in the quantity of interest via the global bilinear form that describes the problem in terms of local and computable quantities.

Our methodology, however, is based on the selection of an alternative bilinear form exhibiting better properties than the original bilinear form (e.g. positive definiteness). We represent the residual error functional of the adjoint problem through this alternative form. We can then compute new upper bounds of the error of the quantity of interest in a similar way than with the classical approach. Our main goal is to demonstrate that a proper choice of such alternative form may improve the upper bounds of the error representation.

Moreover, the method proposed here generalises the existing ones, since, in particular, we can select as the alternative bilinear form the one associated to the adjoint problem.

6.2.2.2. Hybridizable Discontinuous Galerkin method for the elastic Helmholtz equations

Participants: Marie Bonnasse-Gahot, Henri Calandra, Julien Diaz, Stéphane Lanteri.

We consider Discontinuous Galerkin (DG) methods formulated on fully unstructured meshes, which are more convenient than finite difference methods on cartesian grids to handle the topography of the subsurface. DG methods and classical Finite Element (FE) methods mainly differ from discrete functions which are only piecewise continuous in the case of DG approximation. DG methods are then more suitable than Continuous Galerkin (CG) methods to deal with hp-adaptivity. This is a great advantage to DG method which is thus fully adapted to calculations in highly heterogeneous media. Nevertheless, the main drawback of classical DG methods is that they are more expensive in terms of number of unknowns than classical CG methods, especially when arbitrarily high order interpolation of the field components is used. In this case DG methods lead to larger sparse linear systems with a higher number of globally coupled degrees of freedom as compared to CG methods with a same given mesh. In that case, we consider a hybridizable Discontinuous Galerkin (HDG) method which principle consists in introducing a Lagrange multiplier representing the trace of the numerical solution on each face of the mesh cells. This new variable exists only on the faces of the mesh and the unknowns of the problem depend on it. This allows us to reduce the number of unknowns of the global linear system. Now the size of the matrix to be inverted only depends on the number of the faces of the mesh and on the number of the degrees of freedom of each face. It is worth noting that for the classical DG method it depends on the number of the cells of the mesh and on the number of the degrees of freedom of each cell. The solution to the initial problem is then recovered thanks to independent elementwise calculation. The principle of the HDG method and 2D results were presented at the WCCM XI - ECCM V - ECFD VI - Barcelona 2014 Conference [41], the EAGE Workshop on High Performance Computing for Upstream [42], the Second Russian-French Workshop “Computational Geophysics” [43] and at the Réunion des Sciences de la Terre 2014 conference [53]. A comparison between HDG method and classical nodal DG method was given on a poster at the Journées Total-Mathias 2014 workshop [66].

6.2.2.3. Helioseismology

Participants: Hélène Barucq, Juliette Chabassier, Marc Duruflé, Damien Fournier, Laurent Gizon.

The finite element code Montjoie 5.2 has been used to solve Helmholtz equation in axisymmetric domain in the configuration of the sun. The efficiency of the code has been compared in three configurations : radial (1-D mesh and spherical harmonics), axisymmetric (2-D mesh), 3-D. The results have convinced our-selves and our partners of Max Planck Institute that the axisymmetric configuration is the most interesting for an inversion procedure, since 3-D computations are too expensive. A more realistic modeling of the sun requires the solution of time-harmonic Galbrun's equations (instead of Helmholtz equation), different formulations have been implemented and studied. It appeared that the different numerical methods are not able to converge to the correct solution for non-uniform flows. The lack of convergence is more obvious for flows with a larger Mach number. Such problems do not appear in Linearized Euler equations, as a result we have proposed simplified Galbrun's equations that converge correctly and provide the same solution as original Galbrun's equations for a null flow. These equations have been implemented in 2-D, axisymmetric and 3-D configuration.

6.2.2.4. Scattering of acoustic waves by a disc - Hypersingular integral equations

Participants: Leandro Farina, Paul Martin, Victor Péron.

Two-dimensional boundary-value problems involving a Neumann-type boundary condition on a thin plate or crack can often be reduced to one-dimensional hypersingular integral equations. Examples are potential flow past a rigid plate, acoustic scattering by a hard strip, water-wave interaction with thin impermeable barriers, and stress fields around cracks. In [29], we generalize some of these results to two-dimensional hypersingular integral equations. Thus, rather than integrating over a finite interval, we now integrate over a circular disc. Two-dimensional hypersingular equations over a disc arise, for example, in the scattering of acoustic waves by a hard disc; this particular application is described in Appendix A. We develop an appropriate spectral (Galerkin) method, using Fourier expansions in the azimuthal direction and Jacobi polynomials in the radial direction. The Hilbert-space arguments used by Golberg are generalized and a convergence theorem is proved by using tensor-product techniques. Our results are proved in weighted L^2 spaces. Then, Tranter's method is discussed. This method was devised in the 1950s to solve certain pairs of dual integral equations. It is shown that this method is also convergent because it leads to the same algebraic system as the spectral method.

6.2.2.5. Finite Element Subproblem Method

Participants: Patrick Dular, Christophe Geuzaine, Laurent Krähenbühl, Victor Péron.

In the paper [26], the modeling of eddy currents in conductors is split into a sequence of progressive finite element subproblems. The source fields generated by the inductors alone are calculated at first via either the Biot-Savart law or finite elements. The associated reaction fields for each added conductive region, and in return for the source regions themselves when massive, are then calculated with finite element models, possibly with initial perfect conductor and/or impedance boundary conditions to be further corrected. The resulting subproblem method allows efficient solving of parameterized analyses thanks to a proper mesh for each subproblem and the reuse of previous solutions to be locally corrected.

6.2.2.6. High Order Methods for Helmholtz Problems in Highly Heterogeneous Media

Participants: Théophile Chaumont-Frelet, Henri Calandra, Hélène Barucq, Christian Gout.

Heterogeneous Helmholtz problems arise in various geophysical application where they modelize the propagation of time harmonic waves through the subsurface. For example, in inversion problems, the aim is to reconstruct a map of the underground based on surface acquisition. This recovery process involves the solution to several Helmholtz problems set in different media, and high frequency solutions are required to obtain a detailed image of the underground. This observations motivate the design of efficient solver for highly heterogeneous Helmholtz problems at high frequency.

The main issue with the discretization of high frequency problems is the so called “pollution effect” which impose drastic condition on the mesh. In the homogeneous case, it is known that one efficient way to reduce the pollution effect is the use of high order discretization methods. However, high order methods can not be applied as is to highly heterogeneous media. Indeed, they are based on coarser mesh and are not sensitive to fine scale variations of the medium.

We propose to overcome this difficulty by using a multiscale strategy to take into account fine scale heterogeneities on coarse meshes. The method is based on a simple medium approximation method, which can be seen as a special quadrature rule. Numerical experiments in two dimensional geophyscial benchmarks show that high order method coupled with our multiscale approximation medium stragey are cheaper than low order method for a given accuracy. Futhermore, focusing on one dimensional models, we were able to show from a theoretical point of view that our methology reduces the pollution effect even when used on coarse meshes with non-matching interfaces.

This work has been presented at the *WCCM XI - ECCM V - ECFD VI - Barcelona 2014* conference, the *Second Russian-French Workshop "Computational Geophysics"*. A poster has been presented at the *journées Total-Mathias 2014 workshop*. A paper has been submitted for publication to *Math. Of Comp.*.

6.2.3. Boundary conditions.

6.2.3.1. Absorbing Boundary Conditions for Tilted Transverse Isotropic Elastic Media

Participants: Lionel Boillot, Hélène Barucq, Julien Diaz, Henri Calandra.

The seismic imaging simulations are always performed in bounded domains whose external boundary does not have physical meaning. We have thus to couple the wave equations with boundary conditions which aim at reproduce the invisibility of the external boundary. The discretization of these conditions can be an issue. For instance, an efficient condition, once discretized, can induce huge computational costs by filling the matrix which has to be inverted. This is the case of the transparent boundary conditions which are approximated by local Absorbing Boundary Conditions (ABC) that do not increase to much the computational burden. However, the ABC has the drawback to introduce spurious numerical waves which can perturb the RTM results. It is possible to avoid this drawback by applying PML (Perfectly Matched Layers) but it proves to be unstable in anisotropic media. Last year, we proposed a way of construction leading to a stable ABC. The technique is based on slowness curve properties, giving to our approach an original side. We established stability results from long time energy behavior and we have illustrated the performance of the new condition in 2D numerical tests. This year, we extend all these results to 3D case and to arbitrary boundary shapes. The previous paper submission on 2D results has been accepted and released [18]. The recent results in 3D have been presented to the ECCOMAS conference.

6.2.3.2. Derivation of high order absorbing boundary conditions for the Helmholtz equation in 2D.

Participants: Hélène Barucq, Morgane Bergot, Juliette Chabassier, Élodie Estecahandy.

Numerical simulation of wave propagation raises the issue of dealing with outgoing waves. In most of the applications, the physical domain is unbounded and an artificial truncation needs indeed to be carried out for applying numerical methods like finite element approximations. Adapted boundary conditions that avoid the reflection of outgoing waves and provide a well-posed mathematical problem must then be derived. With ideal boundary conditions, the solution on the new mixed boundary valued problem in the truncated domain would actually be equal to the restriction of the mathematical solution in the unbounded domain. However, such ideal boundary conditions, called "transparent boundary conditions", can be shown to be nonlocal, which leads to dramatic computational overcosts. The seek of local boundary conditions, called "absorbing boundary conditions" (ABC), has been the object of numerous works trying to perform efficient conditions based on different techniques of derivation. Among them, the technique of micro-diagonalisation has been employed to the wave equation and more generally to hyperbolic systems in [76], leading to a hierarchy of absorbing local boundary conditions based on the approximation of the Dirichlet-to-Neumann map. A comprehensive review of different used strategies and higher order conditions can be found in [85]. One desirable property of ABCs is that the reflection of the waves on the artificial boundary generates an error of the same order as the one generated by the spatial discretization inside the domain. The computational effort is thus optimized in terms of modeling and numerical inaccuracies. Moreover, the ABC must fit the artificial boundary chosen by the user of the method. In the context of high order spatial discretization (spectral finite elements [74], Interior Penalized Discontinuous Galerkin [68]), there is nowadays a need for high order ABCs that can adapt on non flat geometries since these methods prove very efficient for capturing arbitrary shaped domains.

The aim of the present work is to develop high order ABCs for the Helmholtz equation, that can adapt to regular shaped surfaces. A classical way of designing ABCs is to use Nirenberg theorem [80] on the second order formulation of the Helmholtz equation, which enables us to decompose the operator as a product of two first order operators. Here our approach is to rewrite the Helmholtz equation as a first order system of equations before developing ABCs using M.E. Taylor's micro-diagonalisation method [84]. Then an asymptotic truncation must be performed in order to make the ABC local, and we will see that the high frequency approximation will lead to more usable ABCs than the one stating that the angle of incidence is small. During the process, while increasing the degree of the pseudo differential operator decomposition along with the order of asymptotic truncation, we retrieve classical ABCs that have been found with other techniques by other authors. For now, we have restricted ourselves to two dimensions of space, but despite the fact that 3D generalization should obviously generate more calculation, no further theoretical difficulties are expected.

This work has been the object of a technical report [61] and the obtained conditions have been implemented in Montjoie 5.2 and Houd10ni 5.1 .

6.2.4. Asymptotic modeling.

6.2.4.1. Fast Simulation of Through-casing Resistivity Measurements Using Semi-analytical Asymptotic Models.

Participants: Victor Péron, David Pardo, Aralar Erdozain.

When trying to obtain a better characterization of the Earth's subsurface, it is common to use borehole through-casing resistivity measurements. It is also common for the wells to be surrounded by a metal casing to protect the well and avoid possible collapses. The presence of this metal case highly complicates the numeric simulation of the problem due to the high conductivity of the casing compared to the conductivity of the rock formations. In this study [47] we present an application of some theoretical asymptotic methods in order to deal with complex borehole scenarios like cased wells. The main idea consists in replacing the part of the domain related to the casing by a transmission impedance condition. The small thickness of the casing makes it ideal to apply this kind of mathematical technique. When eliminating the casing from the computational domain, the computational cost of the problems considerably decreases, while the effect of the casing does not disappear due to the impedance transmission conditions. The results show that when applying an order three impedance boundary condition for a simplified domain, it only generates a negligible approximation error, while it considerably reduces the computational cost. For obtaining the numerical results and testing the mathematical models we have developed a Finite Element Code in Matlab. The code works with Lagrange polynomials of any degree as basis functions and triangular shaped elements in two dimensions. The code has been adapted for working with the transmission impedance conditions required by the mathematical models.

6.2.4.2. Modeling the propagation of ultrashort laser pulses in optical fibers.

Participants: Mohamed Andjar, Juliette Chabassier, Marc Duruflé.

In order to model the propagation of an ultrashort laser pulse, the most natural idea is to solve Maxwell's equations in a nonlinear and dispersive medium. Given the considered optical periods (around 10^{-14} seconds), the associated wavelengthes (around 1 millimeter) and the propagation distances (several meters), the direct numerical simulation of these equations by usual numerical techniques (finite elements, explicit time schemes) is impossible because too expensive. The standard procedure is therefore to use approached equations obtained by exploiting legitimate hypotheses in the considered context (slowly varying pulse envelope, narrow spectrum, paraxial approximation ...). These new equations, among them the Nonlinear Schrödinger Equation, are significantly less expensive to solve and we can therefore provide realistic numerical simulations to physicists.

When the pulse propagates in an optical fiber, its spatial profile in the orthogonal plane to the propagation direction in very simple because optical fibers posses a finite (small, often equal to one) number of propagating modes. The equations that originally are stated on a 3D domain can then be written as one spatial dimension equations.

The scientific objective of this internship was to apply the approximation techniques mentioned above in this specific context, in order to obtain one or several equations (depending on the used hypotheses) that model the propagation of ultrashort laser pulses in optical fibers. A matlab code has been developed and integrated in the C++ code Montjoie 5.2 . Numerical simulations have been led in order to observe classical situations of nonlinear fiber optics (Kerr effect, Raman effect, supercontinuum generation, ...).

6.2.4.3. Small heterogeneities in the context of time-domain wave propagation equation : asymptotic analysis and numerical calculation

Participants: Vanessa Mattesi, Sébastien Tordeux.

We have focused our attention on the modeling of heterogeneities which are smaller than the wavelength. The work can be decomposed into two parts : a theoretical one and a numerical one. In the theoretical one, we derive a matched asymptotic expansion composed of a far-field expansion and a near-field expansion. The terms of the far-field expansion are singular solutions of the wave equation whereas the terms of the near-field expansion satisfy quasistatic problems. These expansions are matched in an intermediate region. We justify mathematically this theory by proving error estimates. In the numerical part, we describe the Discontinuous Galerkin method, a local time stepping method and the implementation of the matched asymptotic method. Numerical simulations illustrate these results. Vanessa Mattesi has defended her PhD on this topic[14].

6.2.4.4. Theoretical and numerical investigations of acoustic response of a multiperforated plate for combustion liners

Participants: Vincent Popie, Estelle Piot, Sébastien Tordeux.

Multiperforated plates are used in combustion chambers for film cooling purpose. As the knowledge of the acoustic response of the chamber is essential for preventing combustion instabilities, the acoustic behaviour of the perforated plates has to be modeled. This can be done either by considering the transmission impedance of the plates, or their Rayleigh conductivity.

We have investigated the link between these two quantities thanks to matched asymptotic expansions. Especially the far-field or near-field nature of the physical quantities used in the definition of the impedance and Rayleigh quantity has been enlightened. Direct numerical simulations of the propagation of an acoustic plane wave through a perforated plate are performed and post-treated so that the assumptions underlying the definitions of impedance and Rayleigh conductivity have been checked. The results will be presented at the conference ASME Turbo Expo 2015.

6.3. High Performance methods for solving wave equations

6.3.1. Coupling the DG code with task programming libraries

Participants: Lionel Boillot, Emmanuel Agullo, George Bosilca, Henri Calandra.

The parallelization of the original code is based on a preliminary step of domain decomposition and then on the use of the MPI (Message Passing Interface) library. It is a common choice which works pretty well in most of the classical architectures. However, the parallel efficiency is not optimal and the performance decreases in hybrid architectures. Indeed, we know the number of operations that each sub-domain has to performed but this does not give us the exact time that the computations require. The cluster heterogeneity leads to various automatic optimizations (memory cache, parallel capability, . . .) which are difficult to measure. We have decided to tackle this problem by modifying the parallelism with the use of task programming. We have thus rewritten the DIVA algorithm in a graph of tasks without using the MPI library and we have left to the runtime PaRSEC the choice of when and where to execute each task. The numerical experiments we have performed have confirmed the significant improvement of the parallel efficiency on different architectures like ccNUMA machines or Intel Xeon Phi co-processors. Moreover, the proposed solution is portable on these architectures, this means that none or few modifications are required in the code, allowing to focus on algorithmic aspects in order to preserve the performance. These results have been presented to the EAGE HPC workshop and to the HPCC IEEE conference within a paper have been accepted.

MOISE Project-Team

6. New Results

6.1. Mathematical Modelling of the Ocean Dynamics

6.1.1. Numerical Schemes for Ocean Modelling

Participants: Eric Blayo, Laurent Debreu, Jérémie Demange, Florian Lemarié.

In his PhD, Jérémie Demange has worked on advection-diffusion schemes for ocean models (Supervisors : L. Debreu, P. Marchesiello (IRD)). His work focuses on the link between tracers (temperature and salinity) and momentum advection and diffusion in the non hyperbolic system of equations typically used in ocean models (the so called primitive equations with hydrostatic and Boussinesq assumptions). We also investigated the use of a depth dependent barotropic mode in free surface ocean models. When most ocean models assume that this mode is vertically constant, we have shown that the use of the true barotropic mode, derived from a normal mode decomposition, allows more stability and accuracy in the representation of external gravity waves ([55]). A special focus has also been put on the numerical representation of internal gravity waves (IGW). The normal mode decomposition also allows the computation of IGW characteristic variables and speeds and thus enables the derivation of monotonic advection schemes ([54]).

In 2014, we worked on the stability constraints for oceanic numerical models ([56]). The idea is to carry a deep analysis of these constraints in order to propose new time stepping algorithms for ocean models. Except for vertical diffusion (and possibly the external mode and bottom drag), oceanic models usually rely on explicit time-stepping algorithms subject to Courant-Friedrichs-Lewy (CFL) stability criteria. Implicit methods could be unconditionally stable, but an algebraic system must be solved at each time step and other considerations such as accuracy and efficiency are less straightforward to achieve. Depending on the target application, the process limiting the maximum allowed time-step is generally different. In this paper, we introduce offline diagnostics to predict stability limits associated with internal gravity waves, advection, diffusion, and rotation. This suite of diagnostics is applied to a set of global, regional and coastal numerical simulations with several horizontal/vertical resolutions and different numerical models. We show that, for resolutions finer that 1/2°, models with an Eulerian vertical coordinate are generally constrained by vertical advection in a few hot spots and that numerics must be extremely robust to changes in Courant number. Based on those results, we review the stability and accuracy of existing numerical kernels in vogue in primitive equations oceanic models with a focus on advective processes and the dynamics of internal waves. We emphasize the additional value of studying the numerical kernel of oceanic models in the light of coupled space-time approaches instead of studying the time schemes independently from spatial discretizations. From this study, we suggest some guidelines for the development of temporal schemes in future generation multi-purpose oceanic models.

6.1.2. Coupling Methods for Oceanic and Atmospheric Models

Participants: Eric Blayo, Mehdi-Pierre Daou, Laurent Debreu, Florian Lemarié, Antoine Rousseau.

6.1.2.1. Coupling dimensionally heterogeneous models

The coupling of different types of models is gaining more and more attention recently. This is due, in particular, to the needs of more global models encompassing different disciplines (e.g. multi-physics) and different approaches (e.g. multi-scale, nesting). Also, the possibility to assemble different modeling units inside a friendly modelling software platform is an attractive solution compared to developing more and more complex global models. More specifically one may want to couple 1D to 2D or 3D models, such as Shallow Water and Navier Stokes models: this was the framework of our partnership with EDF, now extended with ARTELIA Group.

Following the work done by Manel Tayachi in her PhD, Medhi Pierre Daou has started implementing and analyzing a coupling between 1D shallow water equations and 3D Navier Stokes equations. In the context of our partnership with ARTELIA, he uses industrial codes (Mascaret, Telemac and OpenFoam). A first implementation has been realized in an academic testcase, and a second one is presently under implementation in a much more realistic context, in the framework of the European project CRISMA.

6.1.2.2. Ocean-atmosphere coupling

Coupling methods routinely used in regional and global climate models do not provide the exact solution to the ocean-atmosphere problem, but an approached one [57]. For the last few years we have been actively working on the analysis of Schwarz waveform relaxation to apply this type of iterative coupling method to air-sea coupling [95], [96], [94]. In the context of the simulation of tropical cyclone, sensitivity tests to the coupling method have been carried out in an ensemblist approach. We showed that with a mathematically consistent coupling, compared to coupling methods en vogue in existing coupled models, the spread of the ensemble is reduced, thus indicating a much reduced uncertainty in the physical solution. In 2014, this work has been the subject of several invited conferences [23], [24], [25], [26] and collaborations with geophysicists [41], [47], [48].

Past year has also been dedicated to the establishment of strong collaborations between the applied mathematics and the climate community to assess the impact of our work on IPCC-like climate models and to go further in the theoretical work by including the formulation of physical parameterizations. As a results, a PhD-thesis (C. Pelletier) funded by Inria has started in fall 2014 in collaboration with the LSCE (Laboratoire des Sciences du Climat et de l'Environnement). Moreover a PPR (*Projet à partenariat renforcé*) called SIMBAD (SIMplified Boundary Atmospheric layer moDel for ocean modeling purposes) is funded by Mercator-Ocean for the next three years. The aim of this project in collaboration with Meteo-France, Ifremer, LMD, and LOCEAN is to derive a metamodel to force high-resolution oceanic operational models for which the use of a full atmospheric model is not possible due to a prohibitive computational cost.

6.2. Development of New Methods for Data Assimilation

6.2.1. Multigrid Methods for Variational Data Assimilation.

Participants: Laurent Debreu, François-Xavier Le Dimet, Arthur Vidard.

In order to lower the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term on the optimal control and the impact of discretization errors on the efficiency of the coarse grid correction step. We show that even if the optimal control problem leads to the solution of an elliptic system, numerical errors introduced by the discretization can alter the success of the multigrid methods. The view of the multigrid iteration as a preconditioner for a Krylov optimization method leads to a more robust algorithm. A scale dependent weighting of the multigrid preconditioner and the usual background error covariance matrix based preconditioner is proposed and brings significant improvements. This work is presented in a paper submitted to QJRMS ([84]). A book chapter on multiresolution methods for data assimilation has also been published ([51]).

6.2.2. Variational Data Assimilation with Control of Model Error

Participant: Arthur Vidard.

One of the main limitations of the current operational variational data assimilation techniques is that they assume the model to be perfect, mainly because of computing cost issues. Numerous researches have been carried out to reduce the cost of controlling model errors by controlling the correction term only in certain privileged directions or by controlling only the systematic and time correlated part of the error.

Both the above methods consider the model errors as a forcing term in the model equations. Trémolet (2006) describes another approach where the full state vector (4D field: 3D spatial + time) is controlled. Because of computing cost one cannot obviously control the model state at each time step. Therefore, the assimilation window is split into sub-windows, and only the initial conditions of each sub-window are controlled, the junctions between each sub-window being penalized. One interesting property is that, in this case, the computation of the gradients, for the different sub-windows, are independent and therefore can be done in parallel.

This method is now implemented in a realistic oceanic framework using OPAVAR/ NEMOVAR. The plan is to extend this study focusing on the parallel aspects of such approach.

6.2.3. Assimilation of Images

Participants: François-Xavier Le Dimet, Maëlle Nodet, Arthur Vidard, Nelson Feyeux, Vincent Chabot, Nicolas Papadakis.

6.2.3.1. Direct assimilation of image sequences

At the present time the observation of Earth from space is done by more than thirty satellites. These platforms provide two kinds of observational information:

- Eulerian information as radiance measurements: the radiative properties of the earth and its fluid envelops. These data can be plugged into numerical models by solving some inverse problems.
- Lagrangian information: the movement of fronts and vortices give information on the dynamics of the fluid. Presently this information is scarcely used in meteorology by following small cumulus clouds and using them as Lagrangian tracers, but the selection of these clouds must be done by hand and the altitude of the selected clouds must be known. This is done by using the temperature of the top of the cloud.

MOISE was the leader of the ANR ADDISA project dedicated to the assimilation of images, and is a member of its follow-up GeoFluids (along with EPI FLUMINANCE and CLIME, and LMD, IFREMER and Météo-France) that just ended in 2013.

During the ADDISA project we developed Direct Image Sequences Assimilation (DISA) and proposed a new scheme for the regularization of optical flow problems [102]. Thanks to the nonlinear brightness assumption, we proposed an algorithm to estimate the motion between two images, based on the minimization of a nonlinear cost function. We proved its efficiency and robustness on simulated and experimental geophysical flows [76]. As part of the ANR project GeoFluids, we are investigating new ways to define distance between a couple of images. One idea is to compare the gradient of the images rather than the actual value of the pixels. This leads to promising results. Another idea, currently under investigation, consists in comparing main structures within each image. This can be done using, for example, a wavelet representation of images. Both approaches have been compared, in particular their relative merits in dealing with observation errors, in a paper accepted late 2014 [8] and presented in several national [37], [38], [27] and international conferences [30], [28], [29].

Vincent Chabot also defended his PhD in July 2014 [2].

In recent developments [17] we have also used "Level Sets" methods to describe the evolution of the images. The advantage of this approach is that it permits, thanks to the level sets function, to consider the images as a state variable of the problem. We have derived an Optimality System including the level sets of the images.

6.2.3.2. Optimal transport for image assimilation

Within the optimal transport project TOMMI funded by the ANR white program (started mid 2011), a new optimization scheme based on proximal splitting method has been proposed to solve the dynamic optimal transport problem. We investigate the use of optimal transport based distances for data assimilation. N. Feyeux started his PhD on this subject last year. The study is still under investigation, but preliminary encouraging results have already been presented twice, in France [68] and Austria [69].

6.2.4. A Nudging-Based Data Assimilation Method: the Back and Forth Nudging

Participants: Maëlle Nodet, Jacques Blum, Didier Auroux.

The Back and Forth Nudging (BFN) algorithm has been recently introduced for simplicity reasons, as it does not require any linearization, nor adjoint equation, or minimization process in comparison with variational schemes. Nevertheless it provides a new estimation of the initial condition at each iteration.

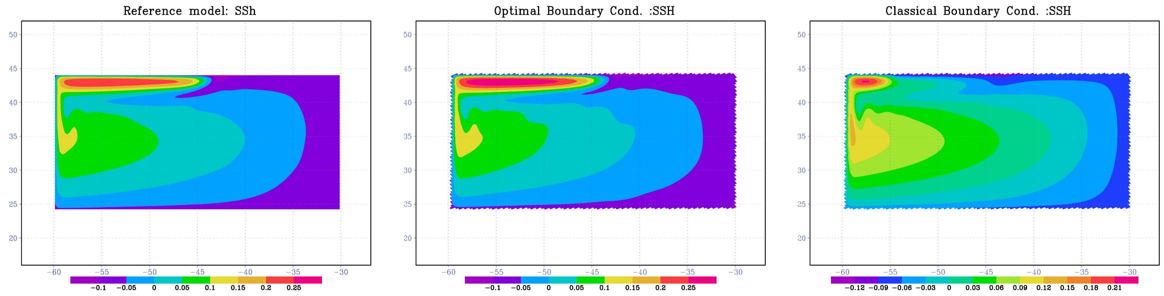
Previous theoretical results showed that BFN was often ill-posed for viscous partial differential equations. To overcome this problem, we proposed a new version of the algorithm, which we called the Diffusive BFN, and which showed very promising results on one-dimensional viscous equations. Experiments on more sophisticated geophysical models, such as Shallow-Water equations and NEMO ocean model are still in progress, in collaboration with University of Nice, and have been presented at the ICIPE conference [31].

6.2.5. Variational Data Assimilation and Control of Boundary Conditions

Participant: Eugène Kazantsev.

A variational data assimilation technique is applied to the identification of the optimal boundary conditions for a simplified configuration of the NEMO model. A rectangular box model placed in mid-latitudes, and subject to the classical single or double gyre wind forcing, is studied. The model grid can be rotated on a desired angle around the center of the rectangle in order to simulate the boundary approximated by a staircase-like coastlines. The solution of the model on the grid aligned with the box borders was used as a reference solution and as artificial observational data. It is shown in [15] that optimal boundary has a rather complicated geometry which is neither a staircase, nor a straight line. The boundary conditions found in the data assimilation procedure bring the solution toward the reference solution allowing to correct the influence of the rotated grid (see fig. 1).

Adjoint models, necessary to variational data assimilation, have been produced by the TAPENADE software, developed by the SCIPORT team. This software is shown to be able to produce the adjoint code, that can be used in data assimilation after a memory usage optimization.



Reference model: SSh

Optimal Boundary Cond. :SSH

Classical Boundary Cond. :SSH

Figure 1. Sea surface elevation: reference solution on the aligned grid (left), solutions on the 30° *rotated grid with optimal (center) and classical (right) boundary conditions.*

6.3. Data Assimilation for Geophysical Models

6.3.1. Development of a Variational Data Assimilation System for OPA9/NEMO

Participants: Arthur Vidard, Pierre-Antoine Bouttier, Eric Blayo.

We are heavily involved in the development of NEMOVAR (Variational assimilation for NEMO). For several years now, we built a working group (coordinated by A. Vidard) in order to bring together various NEMOVAR user-groups with diverse scientific interests (ranging from singular vector and sensitivity studies to specific issues in variational assimilation). It has led to the creation of the VODA (Variational Ocean Data Assimilation for multi scales applications) ANR project (ended in 2012). A new project, part of a larger EU-FP7 project (ERA-CLIM2) has just started in january 2014.
The project aims at delivering a common NEMOVAR platform based on NEMO platform for 3D and 4D variational assimilation. Following 2009-11 VODA activities, a fully parallel version of NEMOTAM (Tangent and Adjoint Model for NEMO) is now available for the community in the standard NEMO version. This version is based on the released 3.4.1 version of NEMO.

We are also investigating variational data assimilation methods applied to high resolution ocean numerical models (see figure 2). This part of the project is now well advanced and encouraging preliminary results are available on an idealised numerical configuration of an oceanic basin. Several novative diagnostics have been also developed in this framework as part of P.A. Bouttier's PhD that was defended early 2014 [1].

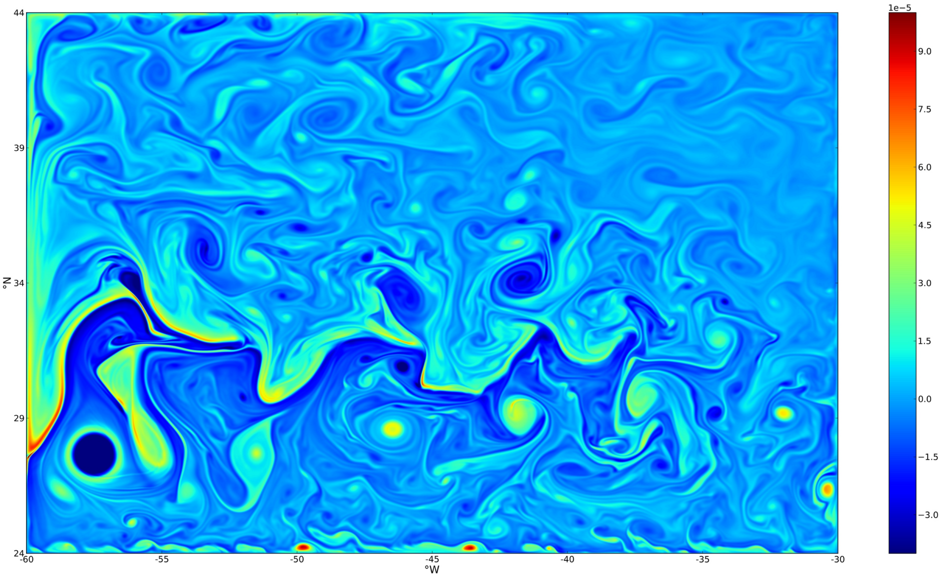


Figure 2. Snapshot of the relative vorticity field (1/s) for an academic oceanic basin model at 1/100° horizontal resolution.

Lastly, multi resolution algorithms have been developed to solve the variational problem. An EU-ITN (International Training Network) project is going to be submitted early 2015 to continue working in this particular aspect.

6.3.2. Ensemble Kalman Filtering for Large Scale Ice-Sheet Models

Participants: Bertrand Bonan, Maëlle Nodet, Catherine Ritz.

In collaboration with C. Ritz (CNRS, Laboratoire de Glaciologie et Geophysique de l'Environnement (LGGE), Grenoble), we aim to develop inverse methods for ice cap models.

In the framework of global warming, the evolution of sea level is a major but ill-known phenomenon. It is difficult to validate the models which are used to predict the sea level elevation, because observations are heterogeneous and sparse.

Data acquisition in polar glaciology is difficult and expensive. Satellite data have a good spatial coverage, but they allow only indirect observation of the interesting data. Moreover, ice dynamics processes are highly non linear and involve many feedback loops, so that classical linear data assimilation gives poor results.

B. Bonan defended his PhD in November 2013 on this subject. We implemented the Ensemble Transform Kalman Filter (ETKF) algorithm for a flowline Shallow-Ice model, called Winnie, developed by C. Ritz at LGGE. On twin experiments we got interesting results, very promising for the future, as we want to implement this method into a full 3D model. A journal paper has published on this subject [5], and the results have been presented in the conference [46].

6.4. Quantifying Uncertainty

6.4.1. Sensitivity Analysis for Forecasting Ocean Models

Participants: Eric Blayo, Laurent Gilquin, Céline Helbert, François-Xavier Le Dimet, Simon Nanty, Maëlle Nodet, Clémentine Prieur, Laurence Viry, Federico Zertuche.

6.4.1.1. Scientific context

Forecasting geophysical systems require complex models, which sometimes need to be coupled, and which make use of data assimilation. The objective of this project is, for a given output of such a system, to identify the most influential parameters, and to evaluate the effect of uncertainty in input parameters on model output. Existing stochastic tools are not well suited for high dimension problems (in particular time-dependent problems), while deterministic tools are fully applicable but only provide limited information. So the challenge is to gather expertise on one hand on numerical approximation and control of Partial Differential Equations, and on the other hand on stochastic methods for sensitivity analysis, in order to develop and design innovative stochastic solutions to study high dimension models and to propose new hybrid approaches combining the stochastic and deterministic methods.

6.4.1.2. Estimating sensitivity indices

A first task is to develop tools for estimated sensitivity indices. In variance-based sensitivity analysis, a classical tool is the method of Sobol' [101] which allows to compute Sobol' indices using Monte Carlo integration. One of the main drawbacks of this approach is that the estimation of Sobol' indices requires the use of several samples. For example, in a d-dimensional space, the estimation of all the first-order Sobol' indices requires $d+1$ samples. Some interesting combinatorial results have been introduced to weaken this defect, in particular by Saltelli [99] and more recently by Owen [97] but the quantities they estimate still require $O(d)$ samples.

In a recent work [21] we introduce a new approach to estimate all first-order Sobol' indices by using only two samples based on replicated latin hypercubes and all second-order Sobol' indices by using only two samples based on replicated randomized orthogonal arrays. We establish theoretical properties of such a method for the first-order Sobol' indices and discuss the generalization to higher-order indices. As an illustration, we propose to apply this new approach to a marine ecosystem model of the Ligurian sea (northwestern Mediterranean) in order to study the relative importance of its several parameters. The calibration process of this kind of chemical simulators is well-known to be quite intricate, and a rigorous and robust — i.e. valid without strong regularity assumptions — sensitivity analysis, as the method of Sobol' provides, could be of great help. The computations are performed by using CIGRI, the middleware used on the grid of the Grenoble University High Performance Computing (HPC) center. We are also applying these estimates to calibrate integrated land use transport models. As for these models, some groups of inputs are correlated, Laurent Gilquin extended the approach based on replicated designs for the estimation of grouped Sobol' indices [70].

We can now wonder what are the asymptotic properties of these new estimators, or also of more classical ones. In [10], the authors deal with asymptotic properties of the estimators. In [89], the authors establish also a multivariate central limit theorem and non asymptotic properties.

6.4.1.3. Intrusive sensitivity analysis, reduced models

Another point developed in the team for sensitivity analysis is model reduction. To be more precise regarding model reduction, the aim is to reduce the number of unknown variables (to be computed by the model), using a well chosen basis. Instead of discretizing the model over a huge grid (with millions of points), the state vector of the model is projected on the subspace spanned by this basis (of a far lesser dimension). The choice of the basis is of course crucial and implies the success or failure of the reduced model. Various model reduction methods offer various choices of basis functions. A well-known method is called “proper orthogonal decomposition" or “principal component analysis". More recent and sophisticated methods also exist and may be studied, depending on the needs raised by the theoretical study. Model reduction is a natural way to overcome difficulties due to huge computational times due to discretizations on fine grids. In [92], the authors present a reduced basis offline/online procedure for viscous Burgers initial boundary value problem, enabling efficient approximate computation of the solutions of this equation for parametrized viscosity and initial and boundary value data. This procedure comes with a fast-evaluated rigorous error bound certifying the approximation procedure. The numerical experiments in the paper show significant computational savings, as well as efficiency of the error bound.

When a metamodel is used (for example reduced basis metamodel, but also kriging, regression, ...) for estimating sensitivity indices by Monte Carlo type estimation, a twofold error appears: a sampling error and a metamodel error. Deriving confidence intervals taking into account these two sources of uncertainties is of great interest. We obtained results particularly well fitted for reduced basis metamodels [14]. In [91], the authors provide asymptotic confidence intervals in the double limit where the sample size goes to infinity and the metamodel converges to the true model. These results were also adapted to problems related to more general models such as Shallow-Water equations, in the context of the control of an open channel [72].

Let us come back to the output of interest. Is it possible to get better error certification when the output is specified. A work in this sense has been submitted, dealing with goal oriented uncertainties assessment [71].

6.4.1.4. Sensitivity analysis with dependent inputs

An important challenge for stochastic sensitivity analysis is to develop methodologies which work for dependent inputs. For the moment, there does not exist conclusive results in that direction. Our aim is to define an analogue of Hoeffding decomposition [90] in the case where input parameters are correlated. Clémentine Prieur supervised Gaëlle Chastaing’s PhD thesis on the topic (defended in September 2013) [78]. We obtained first results [79], deriving a general functional ANOVA for dependent inputs, allowing defining new variance based sensitivity indices for correlated inputs. We then adapted various algorithms for the estimation of these new indices. These algorithms make the assumption that among the potential interactions, only few are significant. Two papers have been recently accepted [66] and [80]. We also considered (see the paragraph 6.4.1) the estimation of groups Sobol’ indices, with a procedure based on replicated designs. These indices provide information at the level of groups, and not at a finer level, but their interpretation is still rigorous.

Céline Helbert and Clémentine Prieur supervise the PhD thesis of Simon Nanty (funded by CEA Cadarache). The subject of the thesis is the analysis of uncertainties for numerical codes with temporal and spatio-temporal input variables, with application to safety and impact calculation studies. This study implies functional dependent inputs. A first step is the modeling of these inputs, and a paper has been submitted [74].

6.4.1.5. Multy-fidelity modeling for risk analysis

Federico Zertuche’s PhD concerns the modeling and prediction of a digital output from a computer code when multiple levels of fidelity of the code are available. A low-fidelity output can be obtained, for example on a coarse mesh. It is cheaper, but also much less accurate than a high-fidelity output obtained on a fine mesh. In this context, we propose new approaches to relieve some restrictive assumptions of existing methods ([93], [98]): a new estimation method of the classical cokriging model when designs are not nested and a nonparametric modeling of the relationship between low-fidelity and high-fidelity levels. The PhD takes place in the REDICE consortium and in close link with industry. The first part of the thesis was also dedicated to the development of a case study in fluid mechanics with CEA in the context of the study of a nuclear reactor.

The second part of the thesis was dedicated to the development of a new sequential approach based on a course to fine wavelets algorithm. Federico Zertuche presented his work at the annual meeting of the GDR Mascot Num in 2014 [36].

6.4.1.6. Data assimilation and second order sensitivity analysis

A main advantage of Variational Methods in Data Assimilation is to exhibit a so-called Optimality System (OS) that contains all the available information : model, data, statistics. Therefore a sensitivity analysis (i.e. the evaluation of the gradient) with respect to the inputs of the model has to be carried out on the OS. With iMECH and INM we have applied sensitivity analysis in the framework of a pollution problem in a lake. The application of second order analysis for sensitivity permits to evaluate the sensitivity with respect to observations and furthermore to determine the optimal location of new sensors at points with the highest sensitivity [16], [52].

This methodology has been applied to

- **Oil Spill**. These last years have known several disasters produced by wrecking of ships and drifting platforms with severe consequences on the physical and biological environments. In order to minimize the impact of these oil spills its necessary to predict the evolution of oil spot. Some basic models are available and some satellites provide images on the evolution of oil spots. Clearly this topic is a combination of the two previous one: data assimilation for pollution and assimilation of images. A theoretical framework has been developed with Dr. Tran Thu Ha (iMech).
- **Data Assimilation in Supercavitation (with iMech)**. Some self propelled submarine devices can reach a high speed thanks to phenomenon of supercavitation: an air bubble is created on the nose of the device and reduces drag forces. Some models of supercavitation already exist but are working on two applications of variational methods to supercavitation:
  - Parameter identification : the models have some parameters that can not be directly measured. From observations we retrieve the unknown parameters using a classical formalism of inverse problems.
  - Shape Optimization. The question is to determine an optimum design of the shape of the engine in order to reach a maximum speed.

6.5. Tracking of Mesoscale Convective Systems

Participant: Clémentine Prieur.

We are interested in the tracking of mesoscale convective systems.A particular region of interest is West Africa. Data and hydrological expertise is provided by T. Vischel and T. Lebel (LTHE, Grenoble).

A first approach involves adapting the multiple hypothesis tracking (MHT) model originally designed by the NCAR (National Centre for Atmospheric Research) for tracking storms [103] to the data for West Africa. With A. Makris (working on a post-doctoral position), we proposed a Bayesian approach [18], which consists in considering that the state at time t is composed on one hand by the events (birth, death, splitting, merging) and on the other hand by the targets' attributes (positions, velocities, sizes, ...). The model decomposes the state into two sub-states: the events and the targets positions/attributes. The events are updated first and are conditioned to the previous targets sub-state. Then given the new events the target substate is updated. A simulation study allowed to verify that this approach improves the frequentist approach by Storlie et al. (2009). It has been tested on simulations [18] and investigated in the specific context of real data on West Africa (submitted paper). Using PHD (probability hypothesis density) filters adapted to our problem, generalising recent developments in particle filtering for spatio-temporal branching processes (e.g. [77]) could be an interesting alternative to explore. The idea of a dynamic, stochastic tracking model should then provide the base for generating rainfall scenarios over a relatively vast area of West Africa in order to identify the main sources of variability in the monsoon phenomenon.

6.6. Multivariate Risk Indicators

Participants: Clémentine Prieur, Patricia Tencaliec.

Studying risks in a spatio-temporal context is a very broad field of research and one that lies at the heart of current concerns at a number of levels (hydrological risk, nuclear risk, financial risk etc.). Stochastic tools for risk analysis must be able to provide a means of determining both the intensity and probability of occurrence of damaging events such as e.g. extreme floods, earthquakes or avalanches. It is important to be able to develop effective methodologies to prevent natural hazards, including e.g. the construction of barrages.

Different risk measures have been proposed in the one-dimensional framework . The most classical ones are the return level (equivalent to the Value at Risk in finance), or the mean excess function (equivalent to the Conditional Tail Expectation CTE). However, most of the time there are multiple risk factors, whose dependence structure has to be taken into account when designing suitable risk estimators. Relatively recent regulation (such as Basel II for banks or Solvency II for insurance) has been a strong driver for the development of realistic spatio-temporal dependence models, as well as for the development of multivariate risk measurements that effectively account for these dependencies.

We refer to [81] for a review of recent extensions of the notion of return level to the multivariate framework. In the context of environmental risk, [100] proposed a generalization of the concept of return period in dimension greater than or equal to two. Michele et al. proposed in a recent study [82] to take into account the duration and not only the intensity of an event for designing what they call the dynamic return period. However, few studies address the issues of statistical inference in the multivariate context. In [86], [88], we proposed non parametric estimators of a multivariate extension of the CTE. As might be expected, the properties of these estimators deteriorate when considering extreme risk levels. In collaboration with Elena Di Bernardino (CNAM, Paris), Clémentine Prieur is working on the extrapolation of the above results to extreme risk levels.

Elena Di Bernardino, Véronique Maume-Deschamps (Univ. Lyon 1) and Clémentine Prieur also derived an estimator for bivariate tail [87]. The study of tail behavior is of great importance to assess risk.

With Anne-Catherine Favre (LTHE, Grenoble), Clémentine Prieur supervises the PhD thesis of Patricia Tencaliec. We are working on risk assessment, concerning flood data for the Durance drainage basin (France). The PhD thesis started in October 2013. A first paper on data reconstruction has been submitted. It was a necessary step as the initial series contained many missing data.

6.7. Non-Parametric Estimation for Kinetic Diffusions

Participant: Clémentine Prieur.

This research is the subject of a collaboration with Venezuela (Professor Jose R. Leon, Caracas Central University) and is partly funded by an ECOS Nord project.

We are focusing our attention on models derived from the linear Fokker-Planck equation. From a probabilistic viewpoint, these models have received particular attention in recent years, since they are a basic example for hypercoercivity. In fact, even though completely degenerated, these models are hypoelliptic and still verify some properties of coercivity, in a broad sense of the word. Such models often appear in the fields of mechanics, finance and even biology. For such models we believe it appropriate to build statistical non-parametric estimation tools. Initial results have been obtained for the estimation of invariant density, in conditions guaranteeing its existence and unicity [6] and when only partial observational data are available. A paper on the non parametric estimation of the drift has been accepted recently [7] (see Samson et al., 2012, for results for parametric models). As far as the estimation of the diffusion term is concerned, a paper has been submitted [7], in collaboration with J.R. León (Caracas, Venezuela) and P. Cattiaux (Toulouse). Recursive estimators have been also proposed by the same authors in [64] recently submitted.

Note that Professor Jose R. León (Caracas, Venezuela) is now funded by an international Inria Chair and will spend one year in our team, allowing to collaborate further on parameter estimation.

6.8. Land Use and Transport Models Calibration

Participants: Thomas Capelle, Laurent Gilquin, Clémentine Prieur, Arthur Vidard, Peter Sturm, Elise Arnaud.

Given the complexity of modern urban areas, designing sustainable policies calls for more than sheer expert knowledge. This is especially true of transport or land use policies, because of the strong interplay between the land use and the transportation systems. Land use and transport integrated (LUTI) modelling offers invaluable analysis tools for planners working on transportation and urban projects. Yet, very few local authorities in charge of planning make use of these strategic models. The explanation lies first in the difficulty to calibrate these models, second in the lack of confidence in their results, which itself stems from the absence of any well-defined validation procedure. Our expertise in such matters will probably be valuable for improving the reliability of these models. To that purpose we participated to the building up of the ANR project CITiES led by the STEEP EPI. This project started early 2013 and two PhD about sensitivity analysis and calibration were launched late 2013. This work led to conference papers [32], [33] and a submitted journal paper [70]

POMDAPI Project-Team

5. New Results

5.1. A posteriori error estimates

Participant: Martin Vohralík.

In [2], we have been able to derive an a posteriori error estimate for the numerical approximation of the two-phase flow problem. This is a cornerstone model problem for porous media, describing the flow of two immiscible and incompressible fluids. We take into account the capillary pressure, whence the model features such difficulties as coupling of partial differential equations with algebraic constraints, strong nonlinearities, degeneracy (disappearance of the diffusion term), advection dominance and consequent forming of sharp evolving fronts, or highly nonlinear and very badly conditioned systems of algebraic equations. Our analysis covers a large class of spatial discretizations in a unified setting, with fully implicit time stepping. We also show how the different error components, namely the spatial discretization error, the temporal discretization error, the linearization error, and the algebraic solver error can be distinguished and estimated separately. This gives rise to efficient adaptive stopping criteria, enabling to spare many useless iterations. The practical impact of our results is that even for this complicated model problem, the overall error committed in a numerical approximation can be fully controlled and, moreover, the simulation time can be reduced by factors typically of an order of magnitude. This result has then been extended in [4] to the compositional model of multiphase Darcy flow, where an arbitrary number of phases can be present, and where each phase can be composed of several components. Later, in [12], still a possible dependence on the temperature has been added. The last two references also contain convincing numerical illustrations on real-life reservoir engineering examples.

5.2. Optimization

Participants: Jean Charles Gilbert, Émilie Joannopoulos, Cédric Josz.

5.2.1. Polynomial optimization

A polynomial optimization problem (POP) consists in minimizing a multivariate real polynomial on a set K defined by polynomial inequalities and equalities. In its full generality it is a non-convex, multi-extremal, difficult global optimization problem. More than a decade ago, J. B. Lasserre proposed to solve a POP by a hierarchy of convex semidefinite programming (SDP) relaxations of increasing size and precision. Each problem in the hierarchy has a primal SDP formulation (a relaxation of a moment expression of the POP) and a dual SDP formulation (a sum-of-squares polynomial relaxation of the POP). In [18], we show that there is no duality gap between each primal and dual SDP problem in Lasserre's hierarchy, provided one of the constraints in the description of set K is a ball constraint. Our proof uses elementary results on SDP duality and it does not assume that K has a strictly feasible point.

5.2.2. Convex quadratic optimization

Convex quadratic optimization deals with problems consisting in minimizing a convex quadratic function on a polyhedron. In [3], we analyzed the behavior of the augmented Lagrangian algorithm when it deals with an *infeasible* convex quadratic optimization problem; this situation is important to master in order to be able to solve correctly the QPs that are generated by the SQP (or Newton-like) algorithm to solve a nonlinear optimization problem, QPs whose feasibility is not guaranteed. It is shown that the algorithm finds a point that, on the one hand, satisfies the constraints shifted by the smallest possible shift that makes them feasible and, on the other hand, minimizes the objective on the corresponding shifted constrained set. The speed of convergence to such a point is globally linear, with a rate that is inversely proportional to the augmentation parameter. This suggests a rule for determining the augmentation parameter that aims at controlling the speed of convergence of the shifted constraint norm to zero; this rule has the advantage of generating bounded augmentation parameters even when the problem is infeasible. The approach has also been implemented in the pieces of software OQLA and QPALM during the ADT MINOQS (see section 4.2 and [16], [14], [15]).

SAGE Project-Team

6. New Results

6.1. Highlights of the Year

Lionel Lenôtre and his co-authors revisited in a very efficient way the Hastings-Metropolis Algorithm on Markov Chains for Small-Probability Estimation.

6.2. Numerical algorithms

6.2.1. Hybrid algebraic sparse linear solvers

Participants: Jocelyne Erhel, David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: [17]

Abstract: Sparse linear systems arise in computational science and engineering. The goal is to reduce the memory requirements and the computational cost, by means of high performance computing algorithms. Krylov methods combined with Domain Decomposition are very efficient. Numerical results show the benefits of our methodology.

6.2.2. GMRES and Polynomial Equivalence

Participant: David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We have established a theoretical link between GMRES and the much simpler problem of polynomial evaluation along with some algebraic structures to describe the most important elements of the GMRES algorithm. We use these structures to show the connection between sequential GMRES and Horner's Rule, s-step GMRES and Dorn's rule, and predict future possible GMRES-like algorithms.

6.2.3. Variables s-step GMRES

Participant: David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We introduce a new variation on s-step GMRES in order to improve its stability, reduce the number of iterations necessary to ensure convergence, and thereby improve parallel performance. In doing so, we develop a new block variant that allows us to express the stability difficulties in s-step GMRES more fully. We use the algebraic structures previous established via the polynomial equivalence to support an intuitive choice for the variation in the s-step procedure, and reinforce its utility in some communication cost estimates.

6.2.4. FGMRES dynamics

Participant: David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: The FGMRES algorithm has met with varying success and we detail theoretical relationships between FGMRES and GMRES including a geometric mean conjecture. Further, we build on the current literature regarding GMRES convergence with an analysis of the dynamical properties of FGMRES.

6.2.5. RPM Coupling Factors

Participant: David Imberti.

Grants and projects: EXA2CT 7.3.1 , C2S@EXA 7.2.3

Publications: in preparation.

Abstract: We have improved the Recursive Projection Method (RPM) with a subspace version that effectively utilizes parallelism. Furthermore, we include a discussion, numerical experiments, and suggestions for the heretofor neglected coupling factor in RPM, and how they influence convergence of the algorithm.

6.2.6. Hastings-Metropolis Algorithm on Markov Chains for Small-Probability Estimation

Participant: Lionel Lenôtre.

Grants: H2MNO4 7.2.1

Publications: [13]

Abstract: Shielding studies in neutron transport, with Monte Carlo codes, yield challenging problems of small-probability estimation. The particularity of these studies is that the small probability to estimate is formulated in terms of the distribution of a Markov chain, instead of that of a random vector in more classical cases. Thus, it is not straightforward to adapt classical statistical methods, for estimating small probabilities involving random vectors, to these neutron-transport problems. A recent interacting-particle method for small-probability estimation, relying on the Hastings-Metropolis algorithm, is presented. It is shown how to adapt the Hastings-Metropolis algorithm when dealing with Markov chains. A convergence result is also shown. Then, the practical implementation of the resulting method for small-probability estimation is treated in details, for a Monte Carlo shielding study. Finally, it is shown, for this study, that the proposed interacting-particle method considerably outperforms a simple Monte Carlo method, when the probability to estimate is small.

6.2.7. A Strategy for the Parallel Implementations of Stochastic Lagrangian Methods

Participant: Lionel Lenôtre.

Grants: H2MNO4 7.2.1

Software: PALMTREE 5.3.1

Publications: [34]

Abstract: We present some investigations on the parallelization of a stochastic Lagrangian simulation. For the self sufficiency of this work, we start by recalling the stochastic methods used to solve Parabolic Partial Differential Equations with a few physical remarks. Then, we exhibit different object-oriented ideas for such methods. In order to clearly illustrate these ideas, we give an overview of the library PALMTREE that we developed. After these considerations, we discuss the importance of the management of random numbers and argue for the choice of a particular strategy. To support our point, we show some numerical experiments of this approach, and display a speedup curve of PALMTREE. Then, we discuss the problem in managing the parallelization scheme. Finally, we analyze the parallelization of hybrid simulation for a system of Partial Differential Equations. We use some works done in hydrogeology to demonstrate the power of such a concept to avoid numerical diffusion in the solution of Fokker-Planck Equations and investigate the problem of parallelizing scheme under the constraint entailed by domain decomposition. We conclude with a presentation of the latest design that was created for PALMTREE and give a sketch of the possible work to get a powerful parallelized scheme.

6.3. Numerical models and simulations applied to physics

6.3.1. Small scale modeling of porous media

Participants: Édouard Canot, Salwa Mansour.

Grants: ARPHYMAT 7.4.3 , 7.4.4

Software: GLiMuH 5.4.3

Publications: [22]

Abstract: This study is devoted to the heat transfer between two spherical grains separated by a small gap; dry air is located around the grains and a liquid water meniscus is supposed to be present between them. This problem can be seen as a micro-scale cell of an assembly of solid grains, for which we are looking for the effective thermal conductivity. For a fixed contact angle and according to the volume of the liquid meniscus, two different shapes are possible for the meniscus, giving a "contacting" state (when the liquid makes a true bridge between the two spheres) and a "non-contacting" one (when the liquid is split in two different drops, separated by a thin air layer); the transition between these two states occurs at different times when increasing or decreasing the liquid volume, thus leading to a hysteresis behavior when computing the thermal flux across the domain.

6.3.2. Heat and mass transfer modeling in porous media

Participants: Édouard Canot, Salwa Mansour.

Grants: HYDRINV 7.4.5

Software: HeMaTiS (5.4.1)

Abstract: The physical model of the HeMaTiS code (5.4.1) has been recently improved by adding the diffusion process of dry air through the water steam which is created by the evaporation of the water inside the porous medium. In this fashion, not only can the heating stage of the surface of the soil be simulated but also the cooling stage. The application concerns the study of archaeological fires which were used many times a day; the possibility of alternation of heating and cooling may lead to a better interpretation of residual marks left in the ground. Work is in progress to validate the numerical results.

6.3.3. Inverse problem for determining the thermo-physical properties of a porous media

Participants: Édouard Canot, Salwa Mansour.

Grants: HYDRINV 7.4.5

Software: TPIP (5.4.2)

Publications: [23]

Abstract: This study concerns the inverse problem which consists of the estimation of thermophysical properties of the soil knowing the temperature at few selected points of the domain. In order to solve this inverse problem, we used the least square criterion where we try to minimize the error function between real measures and simulated ones. The coupled system composed of the energy equation together with the three sensitivity boundary initial problems resulting from differentiating the basic energy equation with respect to the soil properties must be solved. To overcome the stiffness of our problem (due to the use of Apparent Heat Capacity method), the high nonlinearity of the coupled system and the problem of large residuals we used the Damped Gauss Newton and Levenberg-Marquardt methods. Moreover, we emphasized on the importance of the choice of ΔT (temperature range over which the phase change occurs) where for a certain initial guess the inverse problem fails to converge. We overcome this problem by chaining the inverse problems using different values of ΔT and parameters' set.

6.3.4. Geodesy

Participant: Bernard Philippe.

Grants: LIRIMA-EPIC 7.4.2 .

Publications: [12].

Abstract: We solve a geodetic inverse problem for the determination of a distribution of point masses (characterized by their intensities and positions), such that the potential generated by them best approximates a given potential field.

6.4. Models and simulations for flow and transport in porous media

6.4.1. Simulating Diffusion Processes in Discontinuous Media: Benchmark Tests

Participant: Géraldine Pichot.

Grants: H2MN04 7.2.1

Software: SBM 5.2.2

Publications: [33]

Abstract: We present several benchmark tests for Monte Carlo methods for simulating diffusion in one-dimensional discontinuous media, such as the ones arising the geophysics and many other domains. These benchmarks tests are developed according to their physical, statistical, analytic and numerical relevance. We then perform a systematic study on four numerical methods.

6.4.2. Uncertainty Quantification and High Performance Computing for flow and transport in porous media

Participants: Jean-Raynald de Dreuzy, Jocelyne Erhel.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [18]

Abstract: Stochastic models use random fields to represent heterogeneous porous media. Quantities of interest such as macro dispersion are then analyzed from a statistical point of view. In order to get asymptotic values, large scale simulations must be performed, using High Performance Computing. Non-intrusive methods are well-suited for two-level parallelism. Indeed, for each simulation, parallelism is based on domain decomposition for generating the random input and the flow matrix, parallel linear solvers and parallel particle tracker. Also, several simulations, corresponding to randomly drawn samples, can run in parallel. The balance between these two levels depends on the resources available. The software PARADIS implements flow and transport with random data and computation of macro dispersion. Simulations run on supercomputers with large 3D domains.

6.4.3. Computation of macro spreading in 3D porous media with uncertain data

Participants: Jean-Raynald de Dreuzy, Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [15]

Abstract: We consider an heterogeneous porous media where the conductivity is described by probability laws. Thus the velocity, which is solution of the flow equation, is also a random field, taken as input in the transport equation of a solute. The objective is to get statistics about the spreading and the macro dispersion of the solute. We use a mixed finite element method to compute the velocity and a lagrangian method to compute the spreading. Uncertainty is dealt with a classical Monte-Carlo method, which is well-suited for high heterogeneities and small correlation lengths. We give an explicit formulation of the macro dispersion and a priori error estimates. Numerical experiments with large 3D domains are done with the software PARADIS of the platform H2OLab.

6.4.4. A combined collocation and Monte-Carlo method for advection-diffusion equation of a solute in random porous media

Participants: Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [14]

Abstract: In this work, we present a numerical analysis of a method which combines a deterministic and a probabilistic approaches to quantify the migration of a contaminant, under the presence of uncertainty on the permeability of the porous medium. More precisely, we consider the flow equation in a random porous medium coupled with the advection-diffusion equation. Quantities of interest are the mean spread and the mean dispersion of the solute. The means are approximated by a quadrature rule, based on a sparse grid defined by a truncated Karhunen-Loève expansion and a stochastic collocation method. For each grid point, the flow model is solved with a mixed finite element method in the physical space and the advection-diffusion equation is solved with a probabilistic Lagrangian method. The spread and the dispersion are expressed as functions of a stochastic process. A priori error estimates are established on the mean of the spread and the dispersion.

6.4.5. An adaptive sparse grid method for elliptic PDEs with stochastic coefficients

Participants: Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [31].

Abstract: The stochastic collocation method based on the anisotropic sparse grid has become a significant tool to solve partial differential equations with stochastic inputs. The aim is to seek a vector of weights and a convenient level of interpolation for the method. The classical approach uses an a posteriori approach on the solution, which causes an additional prohibitive cost.

In this work, we discuss an adaptive approach of this method to calculate the statistics of the solution. It is based on an adaptive approximation of the *inverse* diffusion parameter. We construct an efficient error indicator which is an upper bound of the error on the solution. In the case of unbounded variables, we use an appropriate error estimation to compute suitable weights for the method. Numerical examples are presented to confirm the efficiency of the approach, and to show that the cost is considerably reduced without loss of accuracy.

6.4.6. Numerical analysis of stochastic advection-diffusion equation via Karhunen-Loève expansion

Participants: Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Publications: [32] , [25]

Abstract: In this work, we present a convergence analysis of a probabilistic approach to quantify the migration of a contaminant, under the presence of uncertainty on the permeability of the porous medium. More precisely, we consider the flow problem in a random porous medium coupled with the advection-diffusion equation and we are interested in the approximation of the mean spread and the dispersion of the solute. The conductivity field is represented by a Karhunen-Loève (K-L) decomposition of its logarithm. The flow model is solved using a mixed finite element method in the physical space. The advection-diffusion equation is computed thanks to a probabilistic Lagrangian method, where the concentration of the solute is the density function of a stochastic process. This process is solution of a stochastic differential equation (SDE), which is discretized using an Euler scheme. Then, the mean of the spread and dispersion are expressed as functions of the approximate stochastic process. A priori error estimates are established on the mean of the spread and of the dispersion. Numerical examples show the effectiveness of this approach.

6.4.7. About a generation of a log-normal correlated field

Participants: Jocelyne Erhel, Mestapha Oumouni.

Grants: HYDRINV 7.4.5 , H2MN04 7.2.1

Software: GENFIELD 5.2.3

Publications: in preparation

Abstract: Uncertainty quantification often requires the generation of large realizations of stationary Gaussian random field over a regular grid.

This paper compares and analyzes the commonalities between the methods and approaches for simulating stationary Gaussian random field. The continuous spectral method is the classical approach which discretizes its spectral density to construct an approximation of the field. When the spectral density and the covariance functions decrease rapidly to zero at infinity, we prove that the spectral method is computationally attractive.

We compare also the classical methods used to simulate the field defined by its covariance function, namely the Discrete Spectral method, the Circulant Embedding approach, and the Discrete Karhunen-Loève approximation. We have found that under some assumptions on the covariance all these latter methods give the same simulations of a stationary Gaussian field on a regular grid, which are very efficient with the Fast Fourier Transform algorithm.

6.4.8. A global model for reactive transport

Participants: Édouard Canot, Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5 , C2SEXA 7.2.3

Software: GRT3D 5.2.1

Thesis: [11]

Publications: [19], [16]

Abstract: In some scientific applications, such as groundwater studies, several processes are represented by coupled models. For example, numerical simulations are essential for studying the fate of contaminants in aquifers, for risk assessment and resources management. Chemical reactions must be coupled with advection and dispersion when modeling the contamination of aquifers. This coupled model combines partial differential equations with algebraic equations, in a so-called PDAE system, which is nonlinear. A classical approach is to follow a method of lines, where space is first discretized, leading to a semi-discrete differential algebraic system (DAE). Several methods have been designed for solving this system of PDAE.

In this study, we propose a global method which uses a DAE solver, where time is discretized by an implicit scheme. Then, each time step involves a nonlinear system of equations, solved by a modified Newton method. Thanks to the DAE solver, the time step is adaptively chosen in order to ensure accuracy and convergence. Moreover, the Jacobian in the nonlinear iterations is freezed as long as Newton converges fast enough, saving a lot of CPU time.

However, the size of the nonlinear system is quite large, because it involves both the differential and the algebraic variables. We show how to eliminate the differential variables, in order to reduce the size. This is equivalent to a so-called Direct Substitution Approach, but it keeps the nice features of DAE solvers.

Classicaly, the concentrations of chemical species are defined with their logarithms, assuming that they are strictly positive. This simplifies the computation of the mass action laws in the chemistry model and the computation of their derivatives. However, when a species does not exist, its concentration is replaced by a very small value and this may lead to an ill-conditioned Jacobian. We propose to use directly the concentrations, without logarithms, so that the Jacobian is then well-conditioned. Therefore, Newton method converges much faster without logarithms, allowing larger time steps and saving many computations.

We illustrate our method with two test cases, provided by the french agency for nuclear waste (ANDRA) and by the group MOMAS. We can compare our results with either analytical or other numerical solutions and show that our method is quite accurate. We also show that reducing the number of unknowns is very efficient and that dealing without logarithms reduces drastically the CPU time.

6.4.9. A chemistry model with precipitation-dissolution

Participant: Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5

Internship: Tangi Migot (Master M2, INSA and University, Rouen)

Publications: [36]

Abstract: In this study, we focus on precipitation and dissolution chemical reactions, because they induce numerical difficulties.

We consider a set of solute species and minerals, with precipitation occuring when a saturation threshold is reached. A challenge is to detect which minerals are dissolved and which minerals are precipitated. This depends on the total quantities of chemical species. We propose an analytical approach to build a phase diagram, which provides the interfaces between the different possible cases. We illustrate our method with three examples arising from brine media and acid mine drainage.

6.4.10. Coupled models for salted aquifers

Participants: Édouard Canot, Jocelyne Erhel.

Grants: H2MN04 7.2.1 , MOMAS 7.2.5 , HYDRINV 7.4.5

Software: GEODENS and SELSAUM (from Tunis)

Internship: Marwen ben Refifa (Ph-D, ENIT, Tunis)

Publications: in preparation

Abstract: We study gravity driven problems in salted aquifers, when many species are present together with high concentrations. In this framework, we couple flow, transport and chemistry by using a fixed point approach. We interfaced two codes developed in Tunis: GEODENS for density driven flow and transport, and SELSAUM for geochemistry. This latter provides also the density of salted water.

6.5. Models and simulations for flow in porous fractured media

6.5.1. A Graphical User Interface for simulating flow and transport in fractured-porous media

Participants: Jean-Raynald de Dreuzy, Jocelyne Erhel, Géraldine Pichot.

Grants: H2MNO4 7.2.1

Platform: H2OLab

Publications: [21]

Abstract: The platform H2OLab can be used with a Graphical User Interface, called H2OGuilde, which is developed using a Qt framework. Launchers correspond to a main program and to a hydrogeological application. These launchers call modules or libraries, implementing discretization schemes, solving algorithms, parallel communications, etc. The interface is generic for all the launchers. It is composed of three main tabs corresponding to the three steps of a simulation: entering input data, running computations, analyzing output data. Input parameters are classified in several categories, corresponding to the physical model and the numerical algorithms chosen. Output parameters are of three types, scalar, vector and matrix. Currently, visualization is done outside of the interface.

6.5.2. Meshing Strategies and the Impact of Finite Element Quality on the Velocity Field in Fractured Media

Participants: Jean-Raynald de Dreuzy, Jocelyne Erhel, Géraldine Pichot.

Grants: FRACINI 7.1.1

Platform: H2OLab

Publications: [20]

Abstract: For solving flow within a network of fractures, Mixed Hybrid Finite Element (MHFE) method is a method of choice as it yields a symmetric positive definite linear system. However, a drawback is its sensitivity to bad aspect ratio elements. For poor-quality triangles, elementary matrices are ill-conditioned and inconsistent velocity vectors are obtained by inverting these local matrices. In our presentation, we will present different strategies for a better reconstruction of the velocity field.

STEEP Team

6. New Results

6.1. Highlights of the Year

This year has seen a number of major advances in the team research projects, on several fronts. The first one concerns the most important and time consuming project, namely integrated land use, activity and transport modelling (LUTI modelling). In this respect, the results described in 6.8 below constitute probably the first set of works contributing sophisticated numerical procedures to the calibration and validation of the TRANUS LUTI model.

The second significant breakthrough concerns the completion of a downscaling methof for Material Flow Analysis (MFA), a key aspect in the characterization and understanding of territorial metabolism for decision-help purposes (section 6.2).

Finally, the modelling effort on land use change for the ESNET project has now been mostly completed, and an operational LUCC model has been calibrated and validated for this project (section 6.3).

6.2. Downscaling Material Flow Analysis: the case of the cereals supply chain in France

The spatial reconstruction of the production, trade, transformation and consumption flows of a specific material can become an important decision-help tool for improving resource management and for studying environmental pressures from the producer's to the consumer's viewpoint. One of the obstacles preventing its actual use in the decision-making process is that building such studies at various geographical scales proves to be costly both in time and manpower. We propose a semi-automatic methodology to overcome this issue. First a supply chain model at the national level has to be designed. Supply and use tables are used to handle the data consistently. Finding the appropriate level of detail for both products and industries is an iterative process: with a small number of highly aggregated product categories, the study isn't likely to provide useful information while with a very detailed list of products and industries, finding input data, especially at local scales, won't be feasible. Secondly, national production, transformation, trade and consumption data have to be reconciled in order to respect the law of mass conservation: this is done through constraint optimization. Thirdly, regional supply and use tables are generated (either with direct data or through downscaling of national data using local proxies, e.g. employment statistics) and reconciled, taking into account the additional constraint that regional data must add up to national one.

We applied the methodology on the case of cereals and reconstructed the supply chain flows of the 22 French regions as well as the flows of four nested territories: France, the Rhône-Alpes région, the Isère département and the territory of the SCOT of Grenoble. Uncertainties of output data were estimated via Monte-Carlo simulations. We display the results using our Sankey diagram vizualisation tool. A research paper is in the reviewing process for one of the major journals in this field.

Future steps include coupling this model with economic (added value), social (local employment) and environmental (environmental pressures) aspects in order to provide new information to decision-makers at various administrative levels (from a group of cities to the national level).

6.3. Mapping and land use and land cover change for the ESNET project

The ESNET project (EcoSystem services NETworks) is a collaboration lead by LECA (Laboratoire d'ECologie Alpine, UJF) that aims at characterizing the ecosystem services of the Grenoble urban region (about 2/3 of the Isere département) at the 2040 horizon under various constraints of urban policy planning, changes in agricultural and forest management, and climate change impact on ecosystems. A preliminary task in this research program was the elaboration of very detailed maps (both in terms of land use and of resolution) of the study area at three different dates (1998, 2003 and 2009) based on available satellite and IGN data, in order to characterize past land use patterns as well as agricultural rotation patterns. These have been made and completed at Inria with the hiring of specialized engineers in these tasks, funded by the ESNET program. This exercice informs the next task (land use and land cover change – LUCC – modelling). Hosting this work at Inria was not only logical in terms of the available computer environment, but also useful in terms of visibility of Inria from outside planning agencies.

The LUCC model itself is developed partly at Inria (for modelling expertise) and partly at LECA (for expertise on ecological change drivers). The model development is now operational, thanks to a major effort on this front in 2014. Both transitions from non urban to urban and use and changes of agricultural practices are now calibrated and validated. The first scenario has been successfully simulated in terms of land use. The three other scenairios of the project are in the final stage of elaboration before simulation, so that the land use change simulation phase of the ESNET project should be completed by the end of April, 2015.

Two resarch papers are in the process of being written on the question of land use practices and their evolution in the study area, and a third one on issues of principle in land use modelling is also underway.

6.4. Benchmarking tools for the climate negotiation of GHG emission reduction trajectories

Climate negotiations related to global warming are another important issue of sustainable development. In this framework that is place at international scale we propose a benchmarking tool that is designed to avoid the main limitations of actual negotiation schemes. Our approach is based on the original Soft Landing proposition, made by Criqui and Kouvaritakis in the early 2000. We develop an up to date solution which improves the original idea mainly by introducing common but differentiated emission reduction profiles and by developing a dedicated algorithm for that purpose (called REDEM). To be compatible with global objectives, it is commonly accepted that for most developing regions, the national emission curves should admit a maximum and then should progressively decline. Similarly, we emphasize the fact that, in order to achieve the global objectives, all states will have to entail mitigation efforts, the intensity which may be measured by the rate of variation of the national emissions. At one point, the effort will reach a maximum, when the rate of variation in absolute value is at its maximum, and then decrease. In other words, there will also be a peak in the effort. Then we propose to base the benchmark on this peak of effort. This work has been done in collaboration with EDDEN Laboratory, in particular Patrick Criqui and Constantin Ilasca. It has been published in [5].

6.5. On the acceptability of land use transport integrated models by French end users as operational tools: from understanding to daily use

Land Use and Transport Integrated models (LUTIs) are promising approaches for urban planning. There is large literature describing their technical architectures or using them in various scientific contexts. Yet little attention has been paid to expectations of practitioners (planners) and to the daily use of such models. There is clearly an important gap between research and practice: a daily use of LUTIs for the simulation of regional planning policies is still an exception in France, despite important research investments and recent interest of planning agencies., and this situation does not seem to be specific to France. We worked on sheding light on what would make them definitely accepted and more used by planners to evaluate a range of urban and transport policies. To do so, we have interviewed different types of end users in France to identify their motivations and barriers to use LUTI models, in addition to literature study and our own experience dealing with urban planning agencies. We have analysed the main obstacles that prevent LUTIs from being widely used by local authorities. It is important to identify that there are two main issues: 1) Do current LUTIs really answer the questions and practical issues territorial agencies are confronted with on a day-to-day basis? Do they match their interests and expectations? 2) Are current LUTIs suitable with respect to the constraints and limitations of local agencies? The main obstacles associated with these issues are: first, it is difficult to match rather generic models with very specific and varied end users questions; second, it is costly and heavy to implement and use a LUTI (capacity obstacles); third, there is no guarantee that results of a dedicated LUTI will have any impact on the policy design (decision making obstacles). The results of our analysis show demand for a far more bottom-up oriented approach: the models should consider objectives and general needs of end users to live up to their expectations. Only a closer collaboration between modelers and end users, and more efforts to integrate modeling into urban planning, will make LUTIs considered as relevant approaches.

This work has been done in collaboration with Mathieu Saujot (IDDRI) and Mathieu De Lapparent (IFSSTAR), and belongs to the work program of CiTIES project.

6.6. Replication procedure for grouped Sobol' indices estimation in dependent uncertainty spaces

Sensitivity analysis studies how the uncertainty on an output of a mathematical model can be attributed to sources of uncertainty among the inputs. Global sensitivity analysis of complex and expensive mathematical models is a common practice to identify influent inputs and detect the potential interactions between them. Among the large number of available approaches, the variance-based method introduced by Sobol' allows to calculate sensitivity indices called Sobol' indices. Each index gives an estimation of the influence of an individual input or a group of inputs. These indices give an estimation of how the output uncertainty can be apportioned to the uncertainty in the inputs. One can distinguish first-order indices that estimate the main effect from each input or group of inputs from higher-order indices that estimate the corresponding order of interactions between inputs. This estimation procedure requires a significant number of model runs, number that has a polynomial growth rate with respect to the input space dimension. This cost can be prohibitive for time consuming models and only a few number of runs is not enough to retrieve accurate informations about the model inputs.

The use of replicated designs to estimate first-order Sobol' indices has the major advantage of reducing drastically the estimation cost as the number of runs becomes independent of the input space dimension. The generalization to closed second-order Sobol' indices relies on the replication of randomized orthogonal arrays. The motivation of this work is to extend this methodology in presence of dependent inputs. Indeed, the case of correlated parameters has to be tackled with caution, as the calculation of single input indices does not provide anymore a proper information, that can be easily interpreted. One strategy is thus to define grouped indices for groups of correlated variables. We address this issue by proposing an approach based on replicated designs and randomized orthogonal arrays that enables to take into account dependency within inputs. We suppose that this dependency can be expressed through constraints. This approach can be used facing any set of constraints at the condition that one is able to provide points in the input space that verify the considered constraints. Guided by our application on a land-use and transport integrated model (LUTI) where some economical parameters are linked by order relations, we focus on the case of sets of linear ordered constraints. Thus we propose a sampling strategy based on the simplex geometric structure, that ensures a proper input space filling.

This work has been done in collaboration with Laurent Gilquin and Clementine Prieur (members of Moise Team), and belongs to the work program of CiTIES project. It is described in [18]. The proposed procedure will be soon applied to study the sensitivity of TRANUS model.

6.7. Specifications for the calibration of Simbad model

"Simbad" is a LUTI model developed by LET. In the context of the CITiESANR project, we have done a comprehensive and detailed study of the parameters of the model in order to fully specify the calibration process of the model. For example, we have specified the objects of interest and indicators, as well as satisfaction criteria. This work has been done in close collaboration with LET.

6.8. Calibration of the TRANUS Land Use Module

The setting up of a LUTI model requires, like most numerical models, at least one phase of parameter estimation. This is concisely referred to here as calibration, although the calibration of a LUTI model also entails other aspects such as the definition of spatial zones, of economic sectors, etc. The TRANUS LUTI model plus software, like many other existing models, come along with a relatively simple calibration methodology. Most LUTI models indeed perform parameter estimation in a piecewise fashion, by sequentially estimating subsets of parameters. While this reduces the mathematical and computational complexity of calibration, neglecting the interactions across different modules and their parameters, may result in a significant loss of a model's quality. A second issue is that TRANUS, like several other LUTI softwares, employs rudimentary numerical routines for parameter estimation. We aim at reducing these weaknesses.

To do so, we first defined a particular parameter estimation problem for TRANUS properly as an optimisation problem, based on an explicit cost function that is to be minimised (something lacking in many articles on LUTI calibration). Next, we developed a series of numerical estimation schemes to solve this optimisation problem. The main difficulty here was that the model is dynamic; by delving into the model's equations and structure, we were able to unwind the model's dynamics and to make it amenable to standard numerical optimisation by gradient descent type methods [4]. This was first done for the estimation of a particular subset of model parameters (the so-called shadow prices). We have recently started to work on the simultaneous estimation of these and other model parameters.

This work is done in collaboration with Arthur Vidard from the MOISE Inria project-team and Brian Morton from the University of North Carolina at Chapel Hill.

6.9. State of the Art on the Calibration and Validation of LUTI Models

One of the tasks of the CITiES project is to construct an extensive state of the art report on the calibration and validation of LUTI models. We coordinate this effort, which involves all partners of CITiES, together with the project partner LVMT (Nicolas Coulombel). It consists of the definition of a taxonomy, of an extensive literature research and of a critical analysis of this literature. A short publication that explains the goals of this effort and some intermediate findings, has been presented in [3]. The completion of this task is expected for the first semester of 2015.

TONUS Team

6. New Results

6.1. Highlights of the Year

We have implemented an OpenCL task graph version of our Discontinuous Galerkin solver that allows to overlap GPU computations and MPI communications. With this optimizations, we were recently able to achieve a 14 GFLOPS simulation with 8 GPUs on an electromagnetic test case. These results are included in the PhD of Thomas Strub (defence planned in March 2015) under the supervision of Philippe Helluy.

6.2. Development of semi-Lagrangian methods

Participants: Adnane Hamiaz, Michel Mehrenberger, Christophe Steiner.

6.2.1. Gyroaverage operator for a polar mesh

A direct method is proposed in [17] in the space configuration for the computation of the gyroaverage operator. It consists in integrating on the gyrocircles using interpolation operators (Hermite or cubic splines); see also [2]. Numerical comparisons with a standard method based on a Padé approximation are performed: (i) with analytical solutions; (ii) considering the 4D drift-kinetic model with one Larmor radius and (iii) on the classical linear DIII-D benchmark case. In particular, we show that in the context of a drift-kinetic simulation, the proposed method has similar computational cost as the standard method and its precision is independent of the radius. Extension to the quasi neutral equation has begun on a 4D model with one Larmor radius. We can exhibit some specific situations where the new method leads to more accurate results and we observe as predicted that the instability growth rate is stronger than for the Padé approximation. On the other hand, we have to face with more oscillations (e.g. on the boundary) of the new operator, which does not permit to replace the Padé approximation. Promising higher order Padé approximation are envisaged for the future.

6.2.2. Semi-Lagrangian simulations on curvilinear grids

Semi-Lagrangian schemes often deal with cartesian mesh; the extension to curvilinear grids is important in order to be able to deal with specific geometries and also for adapting the grid to save computational effort. This study is part of a general work on adding curvilinear capabilities for the simulation of drift kinetic and gyrokinetic equations in a semi-Lagrangian framework, and is in current development in the SeLaLib library.

Thus, in [28] semi-Lagrangian guiding center simulations are performed on sinusoidal perturbations of cartesian grids, thanks to the use of a B-spline finite element solver for the Poisson equation and the classical backward semi-Lagrangian method (BSL) for the advection. We are able to reproduce the standard Kelvin-Helmholtz instability test on such grids. When the perturbation leads to a strong distorted mesh, we observe that the solution differs if one takes standard numerical parameters that are used in the cartesian reference case. We can recover good results together with correct mass conservation, by diminishing the time step.

6.2.3. Field aligned semi-Lagrangian schemes

In [23] we introduce field aligned interpolation for Semi-Lagrangian schemes, by adapting a method developed by Hariri-Ottaviani to the semi-Lagrangian context. This approach is validated on the constant oblique advection equation and on a 4D drift kinetic model with oblique magnetic field in cylindrical geometry. The strength of this method is that one can reduce the number of points in the longitudinal direction. Extension to tokamak conguration in toroidal geometry is the next step of this study.

6.2.4. KEEN wave simulations, high order time splitting, non-uniform cubic splines

KEEN waves are non-stationary, nonlinear, self-organized asymptotic states in Vlasov plasmas (see [3]). They lie outside the precepts of linear theory or perturbative analysis, unlike electron plasma waves or ion acoustic waves. Steady state, nonlinear constructs such as BGK modes also do not apply. The range in velocity that is strongly perturbed by KEEN waves depends on the amplitude and duration of the ponderomotive force generated by two crossing laser beams, for instance, used to drive them. Smaller amplitude drives manage to devolve into multiple highly-localized vorticlets, after the drive is turned off, and may eventually succeed to coalesce into KEEN waves. Fragmentation once the drive stops, and potential eventual remerger, is a hallmark of the weakly driven cases. A fully formed (more strongly driven) KEEN wave has one dominant vortical core. But it also involves fine scale complex dynamics due to shedding and merging of smaller vortical structures with the main one. Shedding and merging of vorticlets are involved in either case, but at different rates and with different relative importance. The narrow velocity range in which one must maintain sufficient resolution in the weakly driven cases, challenges fixed velocity grid numerical schemes. What is needed is the capability of resolving locally in velocity while maintaining a coarse grid outside the highly perturbed region of phase space. We here report on a new Semi-Lagrangian Vlasov-Poisson solver based on conservative non-uniform cubic splines in velocity that tackles this problem head on. An additional feature of our approach is the use of a new high-order time-splitting scheme which allows much longer simulations per computational effort. This is needed for low amplitude runs. There, global coherent structures take a long time to set up, such as KEEN waves, if they do so at all. The new code's performance is compared to uniform grid simulations and the advantages are quantified. The birth pains associated with weakly driven KEEN waves are captured in these simulations. Canonical KEEN waves with ample drive are also treated using these advanced techniques. They will allow the efficient simulation of KEEN waves in multiple dimensions, which will be tackled next, as well as generalizations to Vlasov-Maxwell codes. These are essential for pursuing the impact of KEEN waves in high energy density plasmas and in inertial confinement fusion applications. More generally, one needs a fully-adaptive grid in- phase-space method which could handle all small vorticlet dynamics whether pealing or remerging. Such fully adaptive grids would have to be computed sparsely in order to be viable. This two-velocity grid method is a concrete and fruitful step in that direction.

6.2.5. Conservative semi-Lagrangian scheme

While developing a new semi-Lagrangian solver, the gap between a linear Landau run in 1D-1D and a 5D gyrokinetic simulation in toroidal geometry is quite huge. Intermediate test cases are welcome for testing the code. A new fully two-dimensional conservative semi-Lagrangian (CSL) method is presented in [6] and is validated on 2D polar geometries. We consider here as building block, a 2D guiding-center type equation on an annulus and apply it on two test cases. First, we revisit a 2D test case previously done with a PIC approach and detail the boundary conditions. Second, we consider a 4D drift-kinetic slab simulation. In both cases, the new method appears to be a good alternative to deal with this type of models since it improves the lack of mass conservation of the standard semi-Lagrangian (BSL) method.

6.3. Reduced Vlasov-Maxwell modeling

Participants: Philippe Helluy, Laurent Navoret, Thi Trang Nhung Pham.

We have tested several preliminary methods for reducing the complexity of the Vlasov equation. By expanding the distribution function on velocity basis we obtain a space-only hyperbolic system. This system takes advantage of interesting conservation or entropy properties. Several types of basis can be used: Fourier [14], piecewise Lagrange [20], [13], etc. The method has been implemented for 4D problems in the Selalib library. The next step would be to adapt the size of the expansion according to the nature of the flow region and to apply the method to the gyrokinetic model.

6.4. GPU Optimization of Discontinuous Galerkin solvers

Participants: Michaël Gutnic, Philippe Helluy, Michel Massaro, Thomas Strub.

We have continued to investigate implementations of numerical schemes on new hybrid computer architectures. We have for instance applied a very efficient Strang splitting algorithm for the numerical resolution of the MHD or compressible multiphase model ([16], [18], [22]). We have also highly optimized our DG solver CLAC ([20]) for electromagnetic applications. For instance, we have implemented an OpenCL task graph that allows to overlap GPU computations and MPI communications. With this optimizations, we were recently able to achieve a 14 GFLOPS simulation with 8 GPUs on an electromagnetic test case. These results are included in the PhD of Thomas Strub (defence planned in March 2015).

6.5. Numerical and theoretical study of reduced MHD problems for the JOREK code

Participant: Emmanuel Franck.

The Jorek code is a parallel finit element code (used at the CEA Cadarache and the IPP) which simulates the edge instabilities in the Tokamak solving reduced MHD models. Firstly we have written a family of full MHD models (resistive, diamagnetic and extended MHD models). Using this, we write the reduced MHD models close to the models implemented in the code which conserve the energy and are more stable ([35]). This work will probably be published as an Inria report next year. The second part of this work consists in writing a simplified version of the JOREK code which will be useful to test and validate future numerical research in the JOREK context. Actually we have written a code which solve simple elliptic equations in 3D toroidal geometry using Bezier, splines and Fourier expansion. The integration of simple wave model and reduced MHD models [33] is in progress. When these model will be implemented, we will test a new preconditioning for the JOREK code in these simple configurations.

6.6. Simulations of highly oscillatory Vlasov-type models

Participants: Emmanuel Frénod [Univ. Bretagne-Sud], Sever Hirstoaga.

We continued our exploration of a new time-stepping method based on an exponential integrator.

First, we have improved the algorithm introduced in [11] for solving a multi-scale 1d-1d Vlasov-Poisson system within a Particle-In-Cell method, in order to do accurate long time simulations. As an exponential integrator, the new scheme (see [10]) allows to use large time steps compared to the size of oscillations in the solution. More precisely, the new idea is to push each particle with its computed period. Our simulations show that using precise periods for each particle and at each macroscopic time step results in a more accurate scheme in long times.

Then, similar ideas are used for a 2d-2d multi-scale Vlasov-Poisson system (see [27]). We propose in a Particle-In-Cell framework a robust time-stepping method that works uniformly when the small parameter (the smallest scale) vanishes. We first verify our scheme in the framework of a proposed analytic solution with fast oscillations in time and we show that the scheme works for any initial condition. Then we test the method in the nonlinear case of a Vlasov-Poisson simulation. The scheme is able to use large time steps with respect to the typical size of the solution's fast oscillations. In addition, we show numerically that the method has accurate long time behaviour and that it is asymptotic preserving with respect to the limiting Guiding Center system.

BIOCORE Project-Team

6. New Results

6.1. Highlights of the Year

- We reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976, in a paper published in Ecology Letters [23]. This theorem, also used in human behavior and economics, establishes how individuals should behave to optimize resource exploitation. Despite the thousands of papers written on the subject, we obtained the first mathematical characterization of how habitat characteristics affect the optimal foraging strategy. Mathematical foundations for this work were given in [24].
- The analysis of metabolic networks is generally made under the assumption (so called "balanced growth") that there is no internal accumulation of metabolites. However, this hypothesis is clearly wrong for microalgae, which store lipids and carbohydrates during the day and consume it during the night. A new formalism, called DRUM (Dynamic Reduction of Unbalanced Metabolism) was developed [16], assuming that the balanced growth is valid only in subnetworks, but that there can be accumulation between these modules (which often represent spatial distribution in the cell). This approach was successfully used to represent the dynamics of carbon accumulation in the microalgae *Tisochrysis lutea* under light/dark cycles, or in response to a nitrogen starvation. It also well described the diauxic heterotrophic growth of *Chlorella pyrernoidosa* [11].

6.2. Mathematical methods and methodological approach to biology

6.2.1. Mathematical analysis of biological models

6.2.1.1. Mathematical study of semi-discrete models

Participants: Jean-Luc Gouzé, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Elsa Rousseau, Nicolas Bajeux, Bapan Ghosh.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [99]. We used such models and analyzed their properties in several practical situations that are developed in Section 6.3.2 , some of them requiring such a modeling to describe external perturbations of natural systems, and others to take seasonality into account. External perturbations of interacting populations occur when some individuals are introduced or removed from a natural system, which occurs frequently in pest control applications, either through the direct removal of pests [62], or through the introduction of biological control agents [45], [60], [54]. Seasonality is an important property of most agricultural systems in temperate environments since the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of eco-evolutionary dynamics of plant pathogens [25], [59]

6.2.1.2. Model reduction and sensitivity analysis

Participant: Suzanne Touzeau.

Dynamic models representing complex biological systems with numerous interactions can reach high dimensions and include complex nonlinearities. Especially if data are scarce, identifying the model parameters is then a challenge. So we designed an ad-hoc method based on global sensitivity analysis to simplify the model and determine the most influential parameters. It was applied to a within-host immunological model [30], [61]. This application was part of Natacha Go's PhD thesis, supervised by S. Touzeau and C. Belloc (BioepAR, INRA & Oniris Nantes) [90].

6.2.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Alfonso Carta, Ismail Belgacem, Olivier Bernard, Caroline Baroukh, Jean-Philippe Steyer, Diego de Pereda Sebastian, Francis Mairet.

6.2.2.1. Continuous models analysis

Transcription and translation models in bacteria We study detailed models of transcription and translation for genes in a bacterium, in particular the model of gene expression of RNA polymerase. With techniques of monotone systems, and time scale hypotheses, we can show the stability of the fast part of these systems, and reduce them to much smaller models [49], [48], [47]. We also study other models of the global cellular machinery. This is part of the PhD theses of Ismael Belgacem and Alfonso Carta [12], and done in collaboration with Inria IBIS project-team.

A model of synthesis of a virulence factor In collaboration with J.-A. Sepulchre (INLN Nice), we model the production of a virulence factor by a bacterium in a continuous stirred tank reactor. The production of this enzyme is genetically regulated, and degrades a polymeric external substrate into monomers.

Analysis and reduction of biochemical models In collaboration with D. Ropers (Inria IBIS project team), we address the problem of reduction of large biochemical networks, to decompose the dynamic behavior of the whole system into simpler models. This is the subject of the thesis of S. Casagranda.

Design of a bistable switch to control cellular uptake In joint work with Diego Oyarzún (Imperial College), we explore the idea of constructing a synthetic bistable system using an unbranched metabolic chain with a global enzyme regulator. Bistability can be achieved by choosing an appropriate pattern of regulation and deriving conditions on the promoter dynamic ranges to guarantee a bistable uptake flux. This work started during the visit of Diego to Biocore in October 2014.

Analysis of signaling pathways leading to apoptosis In joint work with Jérémie Roux (Marie Curie Fellow, IRCAN Nice), a cascade of signaling modules leading to apoptosis (or programmed cell death) was implemented and studied through simulations. The goal of this work is to determine whether, and at which stage in the pathway, the system may exhibit bistability. This was the work of Xiao Han's internship.

6.2.2.2. Hybrid models analysis

Piecewise quadratic systems for studying growth rate in bacteria The class of piecewise affine systems was extended to deal with dynamics dependent on dilution due to cell growth rate, leading to switched-piecewise quadratic systems [85]. These new systems use an expression for growth rate that may depend on any number of variables and have several quadratic modes. The behavior of piecewise quadratic systems introduces new features, notably regarding solutions at the thresholds when the vector fields are opposing: not only sliding mode solutions but also oscillatory behavior may happen. Part of this work is in the PhD thesis of Alfonso Carta [12].

Attractor computation using interconnected Boolean networks The method developed in [10] has been extended towards a better characterization of the attractors of the interconnected system in terms of invariant sets [26]. The method was used to test growth rate models in E. Coli using Boolean networks.

Analysis of circadian rhythms in cyanobacteria The model describing the system responsible for the circadian rhythm of cyanobacteria previously proposed in [86] has been improved in [50]. Here, we have tested the robustness of the circadian rhythm with respect to the perturbations inherent to the noisy environment of the cell, including cell growth and division. The interconnection between two models was studied: circadian rhythm and a stochastic model for cell division.

Structure estimation for Boolean models of gene regulation networks The problem of estimating Boolean models of gene networks from few and noisy measurements is addressed in [84], joint work with C. Breindl and F. Allgöwer from the University of Stuttgart. The class of unate or canalizing Boolean functions has been further considered and represented by multi-affine polynomials, leading to a reformulation of the estimation problem as a mixed integer linear program.

Structural principles for the existence of limit cycles in two-dimensional piecewise affine models Using concavity and continuity properties of Poincaré maps, we have derived some structural principles which link the topology of the transition graph to the existence, number and stability of limit cycles in a class of two-dimensional piecewise affine biological models of genetic networks [14].

6.2.2.3. Estimation and control

Optimal allocation of resources in a bacterium We study by techniques of optimal control the optimal allocation between metabolism and gene expression during growth of bacteria [52], in collaboration with Inria IBIS project-team.

Estimation of biological models In a joint work with Diego de Pereda (visiting PhD student), we studied observers and interval observers for models of glucose concentration in diabetes.

6.3. Fields of application

6.3.1. Bioenergy

6.3.1.1. Modelling of microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Philipp Hartmann, Ghjuvan Grimaud, Quentin Béchet, David Demory, Hubert Bonnefond, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae [28] and support model development. These experiments have been carried out in the Lagrangian simulator, under constant or periodic light and temperature, varying the total amount of light dose in the day. The response in terms of storage carbon (triglycerides and carbohydrates) has been observed.

Other experiments were carried out to reproduce the light signal perceived by a cell in a raceway pond [89], derived from hydrodynamical studies [55]. An electronic platform was developed to reproduce this high frequency light signal. The experiments show that the microalgae adapt their pigments to the average light that they have received [28].

The effect in the cell cycle of both the light periodic signal, the temperature and a nitrogen limitation were studied. The strong interactions between the different phases of the cell cycle through checkpoints was highlighted [106]. Temperature turned out to play a key role in modulating metabolic fluxes and synchronization.

The effect of cement flue gas on microalgae growth has been tested. It was demonstrated that this CO_2 source can be used to feed microalgal industrial cultures [114].

Finally a new methodology to measure cell viability has been set up. This approach is very promising to distinguish between net and gross growth rate [22].

These works have been carried out in collaboration with A. Talec, S. Rabouille, E. Pruvost and C. Combe (CNRS/UPMC -Oceanographic Laboratory of Villefranche-sur-Mer).

In collaboration with the IFREMER-PBA team (Nantes) we contributed to a study of the possible associations between microalgae and bacteria to enhance overall productivity [98].

Metabolism of carbon storage and lipid production

A macroscopic model for lipid production by oleaginous microalgae [7] has been previously proposed. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*. It predicts carbohydrate and lipid accumulation, under conditions of light/dark cycles and/or nitrogen deprivation [78], [88], [16].

Modeling the coupling between hydrodynamics and biology

In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on multilayer Saint-Venant system) with microalgae growth was developed [83]. This model is supported by the work of ANGE aiming at reproducing the hydrodynamics of the raceway, with a specific attention to the effect of the paddle wheel on the fluid [55].

Modeling the photosynthesis response to fast fluctuating light

The impact of the hydrodynamics on the light perceived by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [34]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis. A new model accounting for photoacclimation was also proposed [96]. Single cell trajectories were simulated by this software, and the effect on photosynthesis efficiency was assessed using models of photosynthesis [95]. These results were compared to experimental measurements where the high frequency light was reproduced [89].

We also developed a model to reproduce the fluorescence of microalgae during a PAM protocol [51]

Modeling microalgae production processes

The integration of different models developed in the group [81], [101], [7] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth in a photobioreactor under light and nitrogen limitations. The strength of this model is that it takes into account the strong interactions between the biological phenomena (effects of light and nitrogen on growth, photoacclimation ...), temperature effect [82], [111] and the radiative transfer in the culture (light attenuation due to the microalgae).
Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [109]. This model is used to predict lipid production in the perspective of large scale biofuel production. It was also used to assess the potential of France for microalgae, when taking into account the actual 2012 meteorology at the scale of France the use of lands, slope, proximity of nutrients and CO_2 [73].
In the framework of the ANR project Purple Sun, we develop an innovative system for microalgae production: a raceway pond under a greenhouse with semi-transparent photovoltaic panels. To this end, we include in the microalgae model the effect of light wavelength, and we develop a thermic model of the system in order to estimate the culture temperature.

Finally, we provide guidelines for the design of experiments with high informative content that allows an accurate estimation of the parameters concerning the effect of temperature and light on microalgae growth. The optimal experiment design problem was solved as an optimal control problem. E-optimal experiments were obtained by using two discretization approaches namely sequential and simultaneous. Simulation results showed the relevance of determining optimal experimental inputs for achieving an accurate parameter estimation [39].

Nitrogen fixation by nitrogenotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [92]. The main fluxes of carbon and nitrogen are represented in the cell. The accumulation of starch during the day to fuel the nitrogenase working in the absence of oxygen during the night was the key process to explain the nitrogen fixation. The strong influence of the cell cycle was also included in the model. Finally, the model was calibrated and validated with the data of 3 experiments carried out with different duration of the light period and daily dose. The model succeeded to efficiently reproduce the experimental data.

This work is done in collaboration with Sophie Rabouille (CNRS-Oceanographic Laboratory of Villefranche-sur-Mer).

Modeling thermal adaptation in microalgae

We have used the Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. For a constant temperature, we have shown that the optimal temperature trait tends to equal the environment temperature. We then study the case where the temperature is periodically fluctuating [53]. We now use this method at the scale of the global ocean, validating our approach with experimental data sets from 194 species.

Including phytoplankton photoadaptation into biogeochemical models

The complexity of the marine ecosystem models and the representation of biological processes, such as photoadaptation, is very challenging to tackle so that their representation remains an open question. We compared several marine ecosystem models with increasing complexity in the phytoplankton physiology representation in order to assess the consequences of the complexity of photoadaptation models in biogeochemical model predictions. Three models of increasing complexity were considered, and the models were calibrated to reproduce ocean data acquired at the Bermuda Atlantic Time-series Study (BATS) from in situ JGOFS (Joint Global Ocean Flux Study) data. It turns out that the more complex models are trickier to calibrate and that intermediate complexity models, with an adapted calibration procedure, have a better prediction capability [77], [15].

This work is done in collaboration with Sakina Ayata (UPMC-Oceanographic Laboratory of Villefranche-sur-Mer).

6.3.1.2. Control and Optimization of microalgae production

On-line monitoring

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [91]. Several developments were carried out in this direction to improve the design and performances of interval observers, and accounting for a specific structure (*i.e.* triangular) or property (*i.e.* Input to State Stable), [38]. Interval observers were designed for the estimation of the microalgae growth and lipid production within a production process [37] and validated experimentally [36].

Optimization of the bioenergy production systems

Based on simple microalgae models, analytical optimization strategies were proposed. We first focused on the optimal operating conditions for the biomass productivity under day/night cycles using Pontryagin's maximum principle (assuming a periodic working mode) [32].

On the other hand, we assessed strategies for optimal operation in continuous mode using the detailed model for raceways [108], [109]. Two strategies were developed. The first one consists in solving numerically an optimal control problem in which the input flow rate of the raceway is calculated such that the productivity in microalgae biomass is maximized on a finite time horizon. In the second strategy, we translated the optimization problem into a regulation problem. We proposed a simple operational criterion that when integrated in a strategy of closed-loop control allows to attain biomass productivities very near to the maximal productivities obtained with the optimal control. We demonstrated that the practical advantages for real implementation makes our proposed controller a suitable control strategy for optimizing microalgae production in raceways.

We also propose a nonlinear adaptive controller for light-limited microalgae culture, which regulates the light absorption factor (defined by the ratio between the incident light and the light at the bottom of the reactor). We show by numerical simulation that this adaptive controller can be used to obtain near optimal productivity under day-night cycles [103].

Interactions between species

Large scale culture of microalgae for bioenergy involves a large biodiversity (different mutants, invasion, growth-promoting bacteria [98]...). Control of such systems requires to consider the interactions between the different species. Such systems involve bacteria and microalgae, and the competition between these organisms can have several equilibrium points, which can be studied with Monod, Contois and Droop type models [33].

In the framework of the ANR Facteur 4 project, we propose to drive this competition exploring different strategies in order to select species of interest.

We have proposed an adaptive controller which regulates the light at the bottom of the reactor [104]. When applied for a culture with n species, the control law allows the selection of the strain with the maximum growth rate for a given range of light intensity. This is of particular interest for optimizing biomass production as species adapted to high light levels (with low photoinhibition) can be selected.
Strategies to improve the temperature response have been proposed. First we modeled the adaptive dynamics for a population submitted to a variable temperature [53]. This was then used to design experiments aiming at enlarging the thermal niche of a species. Experiments with periodic temperature stresses are currently carried out at the LOV.
Finally, in a more theoretical framework, we studied how to select as fast as possible a given species in a chemostat with two species at the initial instant. Using the Pontryagin maximum principle, we have shown that the optimal strategy is to maintain the substrate concentration to the value maximizing the difference between the growth rates of two species [17]. We now try to extend this result for n species with mutations.

6.3.2. Design of ecologically friendly plant production systems

6.3.2.1. Controlling plant pests

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux, Bapan Ghosh.

Optimization of biological control agent introductions

The question of how many and how frequently natural enemies should be introduced into crops to most efficiently fight a pest species is an important issue of integrated pest management. The topic of natural enemies introductions optimization has been investigated for several years [6] [110], unveiling the crucial influence of within-predator density dependent processes. Since parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, which are prevalent among predatory biocontrol agents,the current modeling effort consists in studying the impact of positive predator-predator interactions on the optimal introduction strategies (PhD of Nicolas Bajeux, [45]). The influence of the spatial structure of the environment on biological control efficacy has also been investigated; first results indicate that spatial structure tends to influence it in quite a same way as intra-specific competition does [60]. An extension of that modeling framework was also studied, that considered state dependent impulsive feedback for the stabilization of a positive equilibrium [54].

Connected research on the influence of space on the establishment capacities of biological control agents is also being pursued both through computer simulations and laboratory experiments on parasitoids of the genus *Trichogramma*. This is the topic of the PhD thesis of Thibaut Morel Journel (UMR ISA); in particular, we show how landscape connectivity or spatial heterogeneity shape establishment dynamics in spatially structured environments [63], [64], [65].

Plant compensation, pest control and plant-pest dynamics

Introducing a plant compartment into our models, we first focused on plant-insect interactions and showed how the level and timing of the pest invasion and pests control interventions could have important effects on the plant's growth pattern and its final biomass. We then modeled plant compensation, which is the process by which some plants respond positively to recover from the effects of pest injury. We have shown that depending on plants and pests characteristics, as well as the level of pest attack, plant overcompensation may or may not happen [35]. Experiments have then been held at UMR ISA on tomato plants facing tutta absoluta invasion; tendencies to compensation have been evidenced, but need to be confirmed through larger scale experiments.

This work is part of the PhD thesis of Audrey Lebon (Cirad), supervised in collaboration with Yves Dumont (Cirad), which has been defended in December 2014.

6.3.2.2. Controlling plant pathogens

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Elsa Rousseau.

Sustainable management of plant resistance

Because in addition to being eaten, plants can also get sick, we studied other forms of biological control dedicated to fight plant pathogens. One such method is the introduction of plant strains that are resistant to one pathogen. This often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. It is therefore necessary to find ways to protect the durability of such resistances, which are a natural exhaustible resource. Experiments were conducted in INRA Avignon, followed by high-throughput sequencing (HTS) to identify the dynamics of several virus strains in competition within host plants. Different plant genotypes were chosen for their contrasted effects on genetic drift and selection they induce on virus populations. Those two evolutionary forces can play a substantial role on the durability of plant resistance. Therefore we fitted a mechanistic-statistical model to these HTS data in order to disentangle the relative role of genetic drift and selection during within-host virus evolution [68], [67]. This is the topic of Elsa Rousseau's PhD thesis, and is done in collaboration with Frédéric Fabre and Benoit Moury (INRA Avignon).

We also represented the pathogen spread in agricultural landscapes [40]. At this scale, we looked at how the landscape structure facilitates or impedes the disease spread among host patches. We showed that, when deploying a host with complete resistance to the pathogen along with a susceptible host, mixed landscapes were always more efficient to hamper the disease spread. However, when using a quantitatively resistant host, aggregating the hosts in different regions could result in a better control of the pathogen spread [41]. This work is part of Julien Papaïx's PhD thesis (MIA, INRA Jouy-en-Josas & BIOGER, INRA Grignon).

Eco-evolutionary dynamics of plant pathogens in seasonal environments

Understanding better pathogen evolution also requires to understand how closely related plant parasites may coexist. Indeed, such coexistence is widespread and is hardly explained through resource specialization. We showed that, in agricultural systems in temperate environments, the seasonal character of agrosystems is an important force promoting evolutionary diversification of plant pathogens [94]. Plant parasites reproduction mode may also strongly interact with seasonality. In this context, we investigated the influence of cyclical parthenogenesis, i.e. the alternation of sexual and asexual reproduction phases, on the eco-evolutionary dynamics of plant parasites [25].

This work was part of the PhD thesis of Magda Castel (Agrocampus Ouest) and has been done in collaboration with Frédéric Hamelin (Agrocampus Ouest), Didier Andrivon (INRA Rennes) and Virginie Ravigné (CIRAD Montpellier).

6.3.3. Biological depollution

6.3.3.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Francis Mairet, Jean-Luc Gouzé.

We have considered the problem of global stabilization of an unstable bioreactor model (e.g. for anaerobic digestion), when the measurements are discrete and in finite number ("quantized"). These measurements define regions in the state space, wherein a constant dilution rate is applied. We show that this quantized control may lead to global stabilization: trajectories have to follow some transitions between the regions, until the final region where they converge toward the reference equilibrium [71].

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [56], we have proposed to define and study the multispecies robustness of bioprocess control laws: given a control law designed for one species, what happens when two or more species are present? We have illustrated our approach with a control law which regulates substrate concentration using measurement of growth activity. Depending on the properties of the additional species, the control law can lead to the correct objective, but also to an undesired monospecies equilibrium point, coexistence, or even a failure point. We now start to develop control laws more robust to the presence of additional species.

Moreno [107] has proposed an optimal strategy for fed-batch bioreactor with substrate inhibition. Thanks to the Pontryagin maximum principle and the Hamilton-Jacobi equation, we have shown that the same strategy is still optimal when mortality is included in the model [79]. We have also studied the problem when the singular arc is non-necessarily admissible everywhere (i.e. the singular control can take values outside the admissible control set). We have pointed out the existence of a frame point on the singular arc above which any singular trajectory is not globally optimal. Moreover, we have provided an explicit way for computing numerically the switching curves and the frame point [46], [19].

6.3.3.2. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Philipp Hartmann, Francis Mairet.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Phycover project is aiming at evaluating the potential of this process [113], [112].

In a first stage, we developed models for anaerobic digestion of microalgae. Two approaches were used: first, a dynamic model has been developed trying to keep a low level of complexity so that it can be mathematically tractable for optimization [100] . On the other hand, we have tested the ability of ADM1 [115] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [102] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

Finally, we have proposed and analysed a three dimensional model which represents the coupling of a culture of microalgae limited by light and an anaerobic digester. We first prove the existence and attraction of periodic solutions. Applying Pontryagin's Maximum Principle, we have characterized optimal controls, including the computation of singular controls, in order to maximize methane production. Finally, we determine numerically optimal trajectories by direct and indirect methods [18].

6.3.3.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE and Pierre Collet (IFPEN).

In the sequel of the pioneering life cycle assessment (LCA) work of [97], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
The improvements due to technological breakthrough (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [27]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [74] under which microalgae are growing.

These studies have allowed to identify the obstacles and limitations which should receive specific research efforts to make this process environmentally sustainable [93].

Finally, some works are aiming at normalizing LCA for microalgae and proposing guidelines to make the LCA more easily comparable [87].

These works have been carried out in collaboration with E. Latrille and B. Sialve (INRA - Laboratory of Environmental Biotechnology, Narbonne).

6.3.4. Models of ecosystems

6.3.4.1. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Optimal foraging and residence times variations

In a pair of papers [23], [24], we reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976. This theorem, also used in human behavior and economics, establishes how individuals should behave to optimize resource exploitation. This result has been has been routinely applied in ecology to understand the foraging strategy of animals such as insect parasitoids used for biological control purposes. We obtained the first mathematical characterization of how habitat characteristics (e.g. patch quality, or the distance between resource patches) affect the optimal foraging strategy. This allowed to confirm or refine MVT predictions, and to provide new predictions in the more realistic case of heterogeneous habitats. Some counterintuitive predictions emerged: making resource patches richer can actually make individuals move more rapidly, contradicting generally admitted earlier predictions.

This work was conducted with Vincent Calcagno (UMR ISA) and Frédéric Hamelin (Agrocampus Ouest).

The handicap paradox

We have continued our investigation of the handicap paradox of sexual selection with the tools of signaling theory. Zahavi's handicap principle, and our game theoretic analysis, explain why an equilibrium displays the "handicap" feature [21]. However, the explanation seems somewhat contrived, so the next question is "how could evolution have reached such a state ?" We have investigated that question with the tools of adaptive dynamics, and reached the conclusion that, if one accepts adaptive dynamics as a model of evolution, and our model of sexual selection, the equilibrium described in our previous article is indeed the limit state of evolution [20].

This work was conducted with Frédéric Hamelin (Agrocampus Ouest).

CARMEN Team

6. New Results

6.1. Highlights of the Year

- New associated team EPICARD (principal investigator N. zemzemi, Y. Coudière and J. Henry). The aim of of this associated team for the first year is to overcome the technical difficulties that we pointed out during the year 2014 in inverse problem for the heart.
- June 2014: Based on a peer-reviewed proposal, the Grand équipement national de calcul intensif (GENCI) has attributed us 3 million core-hours on the national high-performance computing system Turing, to be used in the year 2014.
- December 2014: Based on a peer-reviewed proposal, the Grand équipement national de calcul intensif (GENCI) has attributed us 3.5 million core-hours on the national high-performance computing machines Turing, Curie, and Occigen, to be used in the year 2015.
- LIRYC will fund a 2-year postdoctoral position on simulation of Brugada syndrome, a rare ECG anomaly predictive of sudden cardiac death in young, apparently healthy subjects. This work will be performed in tight collaboration with clinicians at the Haut-Lévèque hospital

6.2. Inverse problem

We tested our method using synthetic data generated with a highly realistic forward model. Propagating action potentials were generated using a monodomain reaction-diffusion model with a Ten Tusscher 2006 membrane model. An anisotropic human heart model at 0.2-mm resolution was used for this purpose. Torso potentials were then computed from the simulated transmembrane currents using a finite-difference torso model at 1-mm resolution with intracavitary blood, anisotropic myocardium, lungs, and an anisotropic skeletal muscle layer. We simulated 20 cases: 5 single stimuli, 1 dual stimulus and 14 re-entry simulations. From the simulated torso potentials a 200-channel body surface map recording was extracted and used to test the inverse methods. Inverse solutions in terms of epicardial potentials were computed both with MFS and with our .new optimal control approach. With our algorithms, we were able to construct the electrical potential on the heart surface with a very good accuracy in terms of correlation coefficient. Thus, we could accurately reconstruct the activation pattern.

6.3. CEPS

- Integration of some ionic models into CEPS (N. Zemzemi and F. Caro). Those developments will allow us now to use CEPS for the benchmark named *Second N-version Cardiac Electrophysiology Benchmark Specification actual developments*, see (benchmarck) for more details.
- Development of usefull tools for the code (compilation in order to check the depandancies, validation and coverage of the code). 16 test cases are now implementated in CEPS. Those test cases concern unit test case as test for small resolution of linear system (for the FE P1 implemented into the code) and validatation test case as the heat PDE problem. Tests verify also the parallel implementation.
- At this time, the development of the bi-domain model in CEPS is in progress in CEPS with N. Zemzemi.
- First integration of the new model of S. Labarthe initiated during his PhD with L. Colin. This task needs improvment for validation in terms of development.

6.4. Numerical Scheme

Y. Coudière, C. Pierre and R. Turpault wrote some new high order FV schemes. The goal of this study is a future implementation in CEPS.

6.5. Mathematical Model

M. Potse, P.E. Becue and F. Caro wrote a new model for numerical simulations for cardiac electrophysiology at the microscopic scale. We interfer with the LIRYC in order to describe, as much as possible, the interactions between the extra-cellular medium and the intra-cellular medium.

DRACULA Project-Team

6. New Results

6.1. Highlights of the Year

- Marine Jacquier and Fabien Crauste (in collaboration with C.O. Soulage and H.A. Soul) published a paper ([18], see also § 6.7) in PLoS ONE 2014.
- Sotiris Prokopiou, Loic Barbarroux, Samuel Bernard, Olivier Gandrillon and Fabien Crauste (in collaboration with J. Mafille, Y. Leverrier, C. Arpin and J. Marvel) published a paper ([21], see also § 6.2) in Computation 2014.
- We organized a session "Deterministic and stochastic models in biology and medicine" at 10th AIMS Conference on Dynamical Systems, Differential Equations and Applications, Madrid (Spain), 7 - 11 July 2014 http://www.aimsciences.org/conferences/2014/.
- Our project entitled "Prion and Alzheimer: mathematical modeling and experiments dealing with a dangerous liaison" has been granted by the French Association France Alzheimer, and has been selected with 3 other projects amongst 14 supported works to be part of a scientific popularizing broadcasting campaign through a short scientific cartoon http://www.francealzheimer.org/projets-soutenus-cette-ann%C3%A9e/lab-alz-comprendre-enjeux-recherche/964 and https://www.youtube.com/watch?v=X0mLf8IJhV4&list=PLCq-e7n2r6Wgo3kaseDHetNAPAG7y9B-d.

6.2. Multi-scale model of the CD8 T cell immune response

We presented in [21] the first multi-scale model of CD8 T cell activation in a lymph node, following an acute infection. CD8 T cell dynamics are described using a cellular Potts model (hence cells are discrete interacting objects), whereas intracellular regulation is associated with a continuous system of nonlinear ordinary differential equations focusing on the dynamics of key proteins. This model allows to reproduce the dynamics of CD8 T cells over a five days period (corresponding to the activation and differentiation into effector cells) and is currently used to characterize the generation of memory cells.

6.3. Mathematical model of hematopoiesis

We investigate in [5] a mathematical model of blood cell production in the bone marrow (hematopoiesis). The model describes both the evolution of primitive hematopoietic stem cells and the maturation of these cells as they differentiate to form the three kinds of progenitors and mature blood cells (red blood cells, white cells and platelets). The three types of progenitors and mature cells are coupled to each other via their common origin in primitive hematopoietic stem cells compartment. The resulting system is composed by eleven age-structured partial differential equations. To analyze this model, we don't take into account cell age-dependence of coefficients, that prevents a usual reduction of the structured system to an unstructured delay differential system. We study the existence of stationary solutions: trivial, axial and positive steady states. Then we give conditions for the local asymptotic stability of the trivial steady state and by using a Lyapunov function, we obtain a sufficient condition for its global asymptotic stability. In some particular cases, we analyze the local asymptotic stability of the positive steady state by using the characteristic equation. Finally, by numerical simulations, we illustrate our results and we show that a change in the duration of cell cycle can cause oscillations.

6.4. The role of spatial organization of cells in erythropoiesis

Erythropoiesis, the process of red blood cell production occurs mainly in the bone marrow. The functional unit of mammalian erythropoiesis, the erythroblastic island, consists of a central macrophage surrounded by adherent erythroid progenitor cells (CFU-E/Pro-EBs) and their differentiating progeny, the erythroblasts. Central macrophages display on their surface or secrete various growth or inhibitory factors that influence the fate of the surrounding erythroid cells. CFU-E/Pro-EBs have three possible fates : a) expansion of their numbers without differentiation, b) differentiation into reticulocytes that are released into the blood, c) death by apoptosis. CFU-E/Pro-EB fate is under the control of a complex molecular network, that is highly dependent upon environmental conditions in the erythroblastic island. In order to assess the functional role of space coupled with the complex network behavior in erythroblastic islands, we developed hybrid discrete-continuous models of erythropoiesis. In [13], a model was developed in which cells are considered as individual physical objects, intracellular regulatory networks are modeled with ordinary differential equations and extracellular concentrations by partial differential equations. We used this model to investigate the impact of an important difference between humans and mice in which mature late-stage erythroblasts produce the most Fas-ligand in humans, whereas early-stage erythroblasts produce the most Fas-ligand in mice. Although the global behaviors of the erythroblastic islands in both species were similar, differences were found, including a relatively slower response time to acute anemia in humans. Also, our modeling approach was very consistent with in vitro culture data, where the central macrophage in reconstituted erythroblastic islands has a strong impact on the dynamics of red blood cell production. Conclusions: The specific spatial organization of erythroblastic islands is key to the normal, stable functioning of mammalian erythropoiesis, both in vitro and in vivo. Our model of a simplified molecular network controlling cell decision provides a realistic functional unit of mammalian erythropoiesis that integrates multiple microenvironmental influences within the erythroblastic island with those of circulating regulators of erythropoiesis, such as EPO and glucocorticosteroids, that are produced at remote sites.

6.5. Mathematical modelling of cell polarization

In [19], a fine description of the behaviour of a nonlinear drift diffusion model inspired from spontaneous cell polarization was performed. This model has Keller Segel type properties and in particular, quantitative proofs were obtained for the convergence to steady state or self similar profile or blow up. The behaviour depends on the mass of the initial data.

6.6. Numerical modelling of cell distribution in blood flow

Properties of blood cells and their interaction determine their distribution in flow. It is observed experimentally that erythrocytes migrate to the ow axis, platelets to the vessel wall, and leucocytes roll along the vessel wall. In [2], a three-dimensional model based on Dissipative Particle Dynamics method and a new hybrid (discrete-continuous) model for blood cells is used to study the interaction of erythrocytes with platelets and leucocytes in flow. Erythrocytes are modelled as elastic highly deformable membranes, while platelets and leucocytes as elastic membranes with their shape close to a sphere. Separation of erythrocytes and platelets in flow is shown for different values of hematocrit. Erythrocyte and platelet distributions are in a good qualitative agreement with the existing experimental results. Migration of leucocyte to the vessel wall and its rolling along the wall is observed.

6.7. Mathematical model of food intake dynamics

In [18], we propose a nonlinear mathematical model of food intake dynamics and body weight dynamics, involving the description of several regulating hormones (leptin, ghrelin, insulin). Using a temporal perturbation of food availability in groups of rats, this model is able to predict body weight and food intake variations by taking into account energy expenditure dynamics based on a memory of the previous food intake. This model also allowed us to estimate the memory lag to approximately 8 days. It also explains how important variations in food availability during periods longer than these 8 days can induce body weight gains.

6.8. Long time existence of weak solutions to cross diffusion models

We pointed out a general entropy structure in cross diffusion systems. We used this structure with duality arguments to build a general framework in which weak solutions exist for a long time. This led to two research articles (one [10] in Siam Journal of Mathematical Analysis and one and [32] recently accepted in Comm. In PDE). This was conducted with the help of the ANR KIBORD.

6.9. Mathematics of Darwin's diagram

Darwin illustrated his theory about emergence and evolution of biological species with a diagram. It shows how species exist, evolve, appear and disappear. Our goal in [8] is to give a mathematical interpretation of this diagram and to show how it can be reproduced in mathematical models. It appears that conventional models in population dynamics are not sufficient, and we introduce a number of new models which take into account local, nonlocal and global consumption of resources, and models with space and time dependent coefficients.

6.10. A micellar on-pathway intermediate step explains the kinetics of prion amyloid formation

In [16], we used a strong interdisciplinary collaboration between mathematicians and biologists to exhibit a new element taking an important role in the development of the pathological prion formation. Indeed, in a previous work by Alvarez-Martinez et al. (2011), the authors pointed out some fallacies in the mainstream interpretation of the prion amyloid formation. It appeared necessary to propose an original hypothesis able to reconcile the in vitro data with the predictions of a mathematical model describing the problem. Here, a model is developed accordingly with the hypothesis that an intermediate on-pathway leads to the conformation of the prion protein into an amyloid competent isoform thanks to a structure, called micelles, formed from hydrodynamic interaction. The authors also compared data to the prediction of their model and proposed a new hypothesis for the formation of infectious prion amyloids.

M3DISIM Team

6. New Results

6.1. Highlights of the Year

- Radomir Chabiniok recruited in starting research position (start Febr 2015);
- PhD Defense of Annabelle Collin;
- "Usine Nouvelle" article.

6.2. Modeling

6.2.1. Mechanics of collective unfolding

Participants: Matthieu Caruel [correspondant], Jean-Marc Allain [LMS], Lev Truskinovsky [LMS].

Mechanically induced unfolding of passive crosslinkers is a fundamental biological phenomenon encountered across the scales from individual macro-molecules to cytoskeletal actin networks. In this work we study a conceptual model of athermal load-induced unfolding and use a minimalistic setting allowing one to emphasize the role of long-range interactions while maintaining full analytical transparency. Our model can be viewed as a description of a parallel bundle of N bistable units confined between two shared rigid backbones that are loaded through a series spring. We show that the ground states in this model correspond to synchronized, single phase configurations where all individual units are either folded or unfolded. We then study the fine structure of the wiggly energy landscape along the reaction coordinate linking the two coherent states and describing the optimal mechanism of cooperative unfolding. Quite remarkably, our study shows the fundamental difference in the size and structure of the folding-unfolding energy barriers in the hard (fixed displacements) and soft (fixed forces) loading devices which persists in the continuum limit. We argue that both, the synchronization and the non-equivalence of the mechanical responses in hard and soft devices, have their origin in the dominance of long-range interactions. We then apply our minimal model to skeletal muscles where the power-stroke in acto-myosin crossbridges can be interpreted as passive folding. A quantitative analysis of the muscle model shows that the relative rigidity of myosin backbone provides the long-range interaction mechanism allowing the system to effectively synchronize the power-stroke in individual crossbridges even in the presence of thermal fluctuations. In view of the prototypical nature of the proposed model, our general conclusions pertain to a variety of other biological systems where elastic interactions are mediated by effective backbones.

6.2.2. Thermodynamical framework for modeling chemical-mechanical coupling in muscle contraction - Formulation and validation

Participants: Matthieu Caruel, Dominique Chapelle [correspondant], Philippe Moireau.

Muscle contraction occurs at the nanoscale of a hierarchical multi-scale structure with the attachment of so-called cross-bridges within sarcomeres, namely, the creation of chemical bonds between myosin heads and specific sites on actin filaments. A cross-bridge in itself can be seen as a special chemical entity having internal mechanical variables - or degrees of freedom - pertaining to the actual geometric configuration, which implies that the free energy of the cross-bridge - whether in an attached or unattached state - must be made dependent on these internal variables (T.L. Hill, Free Energy Transduction And Biochemical Cycle Kinetics, Dover, 2004). This provides a thermodynamical basis for modeling the complex interplay of chemical and mechanical phenomena at the sarcomere level. Within this framework we propose a muscle model with two mechanical variables associated with a cross-bridge. For the action of individual cross-bridges occurring at the nanometer scale, the energy provided by the Langevin thermostat cannot be neglected, and we therefore propose to endow the internal mechanical variables with stochastic dynamics. Important motivations for this modeling choice include the ability to represent (i) the so-called power-stroke phenomenon and (ii) short-time responses of a muscle, e.g. to load steps. Our approach allows for systematic treatment of the model energetics, and in particular one goal of the proposed description is to investigate the potential benefit in mechanical efficiency with systems including - in addition to chemically-induced transformations - thermally-induced conformational changes such as the power-stroke.

6.2.3. Mechanical modeling and numerical methods for poromechanics: Applications to cardiac perfusion

Participants: Bruno Burtschell, Dominique Chapelle [correspondant], Philippe Moireau.

We have previously formulated a rather general modeling framework of poromechanics – formulations that combine solid and fluid components to represent the behavior of a porous medium – to take into account large deformations and rapid fluid flows, see [16]. This allows to consider, in particular, the application of blood perfusion within the cardiac tissue, which features these specific complex phenomena, out of the scope of classical poromechanical models. One of our major objectives now, within the PhD of Bruno Burtschell, is to propose and assess some associated relevant numerical schemes.

Some existing algorithms of fluid-structure interaction, with which our poromechanics formulations feature deep similarities, have been implemented – in FreeFEM++, both in axisymmetric configuration and in 3D – and compared. Their numerical and theoretical analysis – consistency, convergence – has been performed. Then, the adaptation of these algorithms to our poromechanics formulations enabled us to propose a time discretisation well-fitted to our framework, and to present its energy stability analysis. Further perspectives include implementation and numerical validation of this scheme, including special care regarding space discretisation, then integration into FELISCE ("HappyHeart" module).

6.2.4. Personalized modeling for cardiac amyloidosis diagnosis

Participants: Alessandro Felder, Dominique Chapelle, Philippe Moireau, Jean-François Deux [Hôpital Henri Mondor], Thibault Damy [Hôpital Henri Mondor].

Cardiac amyloidosis is a condition induced by pathological deposition of amyloid proteins within muscle tissue and nerves, thus severely impairing the cardiac function and often requiring cardiac transplantation as the only available treatment. Our objective here in a first stage is to use our previously developed patient-specific modeling methodologies to analyse some clinical cases – based on actual patient data – to better apprehend the impact of the pathology on biomechanical properties. Further perspectives include the modeling of the protein deposition and associated tissue remodeling in order to predict the disease evolution in a patient-specific context. This work is performed in collaboration with medical doctors from Hôpital Henri Mondor (Créteil).

6.3. Numerical Analysis

6.3.1. Fourth-order energy-preserving locally implicit discretization for linear wave equations

Participants: Juliette Chabassier [Magique-3d], Sébastien Imperiale [correspondant].

A family of fourth-order coupled implicit-explicit time schemes has been developed. The spatial coupling is done at the boundaries of several non conforming meshes of regions in which we want to simulate propagating waves. A global discrete energy is shown to be preserved and leads to global fourth-order consistency. Numerical results in 1D and 2D have been produced to illustrate the good behavior of the schemes and their potential for the simulation of realistic highly heterogeneous media and strongly refined geometries, for which using an explicit scheme everywhere can be extremely penalizing. Accuracy up to fourth order reduces the numerical dispersion inherent to implicit methods used with a large time step, and makes this family of schemes attractive compared to second order accurate methods in time.

6.4. Model-Data Interaction

6.4.1. A Luenberger observer for reaction-diffusion models with front position data

Participants: Dominique Chapelle, Annabelle Collin, Philipe Moireau [correspondant].

We propose a Luenberger observer for reaction-diffusion models with propagating front features, and for data associated with the location of the front over time. Such models are considered in various application fields, such as electrophysiology, wild-land fire propagation and tumor growth modeling. Drawing our inspiration from image processing methods, we start by proposing an observer for the eikonal-curvature equation that can be derived from the reaction-diffusion model by an asymptotic expansion. We then carry over this observer to the underlying reaction-diffusion equation by an "inverse asymptotic analysis", and we show that the associated correction in the dynamics has a stabilizing effect for the linearized estimation error. We also discuss the extension to joint state-parameter estimation by using the earlier-proposed ROUKF strategy. We then illustrate and assess our proposed observer method with test problems pertaining to electrophysiology modeling, including with a realistic model of cardiac atria. Our numerical trials show that state estimation is directly very effective with the proposed Luenberger observer, while specific strategies are needed to accurately perform parameter estimation – as is usual with Kalman filtering used in a nonlinear setting – and we demonstrate two such successful strategies.

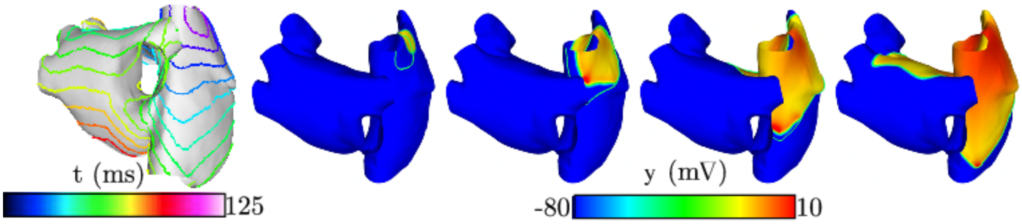


Figure 1. Collocated front data on an atria (left), and observer of the atrial electric activation pursuing the green front from a wrong initial condition (right, 4 time-steps)

6.4.2. Identification of weakly coupled multiphysics problems. Application to the inverse problem of electrocardiography

Participants: Cesare Corrado [Reo team], Jean-Frédéric Gerbeau [Reo team], Philippe Moireau [correspondant].

This work addresses the inverse problem of electrocardiography from a new perspective, by combining electrical and mechanical measurements. Our strategy relies on the definition of a model of the electromechanical contraction which is registered on ECG data but also on measured mechanical displacements of the heart tissue typically extracted from medical images. In this respect, we establish in this work the convergence of a sequential estimator which combines for such coupled problems various state of the art sequential data assimilation methods in a unified consistent and efficient framework. Indeed, we aggregate a Luenberger observer for the mechanical state and a Reduced-Order Unscented Kalman Filter applied on the parameters to be identified and a POD projection of the electrical state. Then using synthetic data we show the benefits of our approach for the estimation of the electrical state of the ventricles along the heart beat compared with more classical strategies which only consider an electrophysiological model with ECG measurements. Our numerical results actually show that the mechanical measurements improve the identifiability of the electrical problem allowing to reconstruct the electrical state of the coupled system more precisely. Therefore, this work is intended to be a first proof of concept, with theoretical justifications and numerical investigations, of the advantage of using available multi-modal observations for the estimation and identification of an electromechanical model of the heart.

6.4.3. Data assimilation for hyperbolic conservation laws. A Luenberger observer approach based on a kinetic description

Participants: Anne-Céline Boulanger [Ange team], Benoît Perthame [Mamba team], Philippe Moireau [correspondant], Jacques Sainte-Marie [Ange team].

Developing robust data assimilation methods for hyperbolic conservation laws is a challenging subject. Those PDEs indeed show no dissipation effects and the input of additional information in the model equations may introduce errors that propagate and create shocks. We propose a new approach based on the kinetic description of the conservation law. A kinetic equation is a first order partial differential equation in which the advection velocity is a free variable. In certain cases, it is possible to prove that the nonlinear conservation law is equivalent to a linear kinetic equation. Hence, data assimilation is carried out at the kinetic level, using a Luenberger observer also known as the nudging strategy in data assimilation. Assimilation then amounts to the handling of a BGK type equation. The advantage of this framework is that we deal with a single "linear" equation instead of a nonlinear system and it is easy to recover the macroscopic variables. The study is divided into several steps and essentially based on functional analysis techniques. First we prove the convergence of the model towards the data in case of complete observations in space and time. Second, we analyze the case of partial and noisy observations. To conclude, we validate our method with numerical results on Burgers equation and emphasize the advantages of this method with the more complex Saint-Venant system.

6.4.4. Optimal observer for parabolic problems

Participants: Karine Mauffrey, Philippe Moireau [correspondant].

We aim at proposing optimal observers strategies for reconstructing the solution of general systems of PDEs using available observations, including both wave-type equations and heat-like equations. The main objective of this work is to present a self-contained analysis. For a general parabolic system, we have established the exponential stability of the operator occurring in the equation satisfied by the error between the target and the optimal observer. The proof relies on two major hypotheses: an observability inequality satisfied by the observation operator and a controllability property for the modeling error operator by which model noises enter the dynamics (controllability property which is related to the invertibility of the solution of the associated infinite dimensional Riccati equation). Then we have tackled the discretisation questions and demonstrated that the discrete-time Kalman filter is an adequate discretization of the continuous-time Kalman filter. Finally we have also studied the strong formulation of the Kalman observer using a kernel representation of the Riccati operator.

6.4.5. Elastography by magnetic resonance imaging

Participants: Guillaume Bal [Columbia Unviersity], Cedric Bellis [LMA Marseille], Sébastien Imperiale [correspondant], Francois Monard [University of Washington- Seattle].

We have studied the potential application of elastography by Magnetic Resonance Imaging (MRI) within the framework of linear elasticity. We assume given internal full-field MRI measurements of the deformations of a non-homogeneous isotropic solid, and the aim is the quantitative reconstruction of the associated physical parameters. Upon using polluted measurements, a variational formulation is constructed, its inversion enabling the recovery of the parameters. The analysis of this inversion procedure provides existence and uniqueness results while the reconstruction stability with respect to the measurements is investigated. As the inversion procedure requires differentiating the measurements twice, a numerical differentiation scheme has been proposed and analyzed. It is based on a regularization that allows an optimally stable reconstruction of the sought parameters.

MAMBA Team

6. New Results

6.1. Highlights of the Year

Benoît Perthame was invited as plenary speaker for the International Congress of Mathematicians ICM 2014 (Seoul, http://www.icm2014.org), that attracted more than 5000 participants. This is the first time that a mathematician working in mathematics applied to biology was invited at ICM, which is the most prestigious conference for mathematicians of all fields. This represents a consecration both for Benoît Perthame's work and for the MAMBA team, and more generally for the whole domain of mathematics applied to biology.

Marie Doumic was a plenary speaker at the ECMTB 2014 (Göteborg, http://ecmtb2014.org/ 600 participants).

Dirk Drasdo was invited speaker at the Systems Biology of Human Diseases conference (Harvard University, http://www.csb2.org/events/sbhd-2014).

Five articles are noteworthy in terms of bibliometry:

• (*Impact factor 11.2*) F. SCHLIESS, S. HOEHME, S. HENKEL, A. GHALLAB, D. DRIESCH, J. BÖTTGER, R. GUTHKE, M. PFAFF, J. HENGSTLER, R. GEBHARDT, D. HÄUSSINGER, D. DRASDO, S. ZELLMER. Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration, *Hepatology*, Dec. 2014, vol. 60, no 6, pp. 2040-2051, https://hal.inria.fr/hal-01110646 [17]

• (*Impact factor 10.4*) D. DRASDO, S. HOEHME, J. G. HENGSTLER. How predictive quantitative modeling of tissue organization can inform liver disease pathogenesis, *Journal of Hepatology*, Oct. 2014, vol. 61, no 4, pp. 951-956 [DOI : 10.1016/J.JHEP.2014.06.013], https://hal.inria.fr/hal-01110644 [7]

• (*Impact factor 10.7*) S.R.K. VEDULA, G. PEYRET, I. CHEDDADI, T. CHEN, A. BRUGUÉS, H. HIRATA, H. LOPEZ-MENENDEZ, Y. TOYAMA, L. NEVES DE ALMEIDA, X. TREPAT, C.T. LIM, B. LADOUX. Mechanics of epithelial closure over non-adherent environments, *Nature Communications*, Jan. 2015, vol. 6, art. number 6111[DOI : 10.1038/ncomms7111], http://www.nature.com/ncomms/2015/150122/ncomms7111/abs/ncomms7111.html (open access)

• (*Impact factor 7.5*) L. ROBERT, M. HOFFMANN, N. KRELL, S. AYMERICH, J. ROBERT, M. DOUMIC. Division in Escherichia coli is triggered by a size-sensing rather than a timing mechanism, in "BMC Biology", 2014, vol. 12, no 1, 17 p. [DOI : 10.1186/1741-7007-12-17], https://hal.inria.fr/hal-00981312 [16]

• (*Impact factor 9.3*) R. H. CHISHOLM, T. LORENZI, A. LORZ, A. K. LARSEN, L. ALMEIDA, A. ESCARGUEIL, J. CLAIRAMBAULT. Emergence of drug tolerance in cancer cell populations: an evolutionary outcome of selection, nongenetic instability and stress-induced adaptation, *Cancer Research* (Mathematical oncology), 10p.+suppl. mat., in press, Jan. 2015, https://hal.archives-ouvertes.fr/hal-01111271 [33]

6.2. Cancer

Participants: Luís Lopes Neves de Almeida, José Luis Avila Alonso [DISCO Inria team], Catherine Bonnet [DISCO Inria team], Rebecca Chisholm, Jean Clairambault, François Delhommeau [Hæmatology department, St Antoine Hospital, Paris], Luna Dimitrio [former PhD student and Mamba member], Ján Eliaš, Alexandre Escargueil [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Pierre Hirsch [Hæmatology department, St Antoine Hospital, Paris], Michal Kowalczyk [Univ. Santiago de Chile], Annette Larsen [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Tommaso Lorenzi, Alexander Lorz, Anna Marciniak-Czochra [Univ. Heidelberg], Roberto Natalini [IAC-CNR, Univ. Tor Vergata, Rome], Silviu Iulan Niculescu [DISCO Inria team], Hitay Özbay [Bilkent Univ., Ankara], Benoît Perthame, Andrada Maran, Fernando Quirós [Univ. Autónoma de Madrid], Michèle Sabbah [Cancer biology and therapeutics lab, St Antoine Hospital, Paris], Thomas Stiehl [Univ. Heidelberg], Min Tang [Jiaotong University, Shanghai], Emmanuel Trélat [LJLL, UPMC], Nicolas Vauchelet, Romain Yvinec [INRA Tours].

6.2.1. Drug resistance.

We have continued to develop our phenotypically based models of drug-induced drug resistance in cancer cell populations, representing their Darwinian evolution under drug pressure by integro-differential equations. In one of them [40], a 1D space variable has been added to the phenotypic structure variable to account for drug diffusion in tumour spheroids. In another one [33], where deterministic and agent-based modelling are processed in parallel, we have considered a physiologically based 2-dimensional phenotypic structure variable, in order to take account of previously published biological observations on (reversible) drug tolerance persistence in a population of non-small cell lung cancer (NSCLC) cells [0], reproducing the observations and assessing the model by testing biologically based hypotheses. Together with ongoing work with E. Trélat and A. Lorz on drug therapy optimisation, using such phenotype-based models to overcome drug resistance, this has represented a significant part of our work on the subject, which is conducted in close collaboration with the INSERM-UPMC team "Cancer biology and therapeutics" (A. Larsen, A. Escargueil, M. Sabbah) at St Antoine Hospital.

6.2.2. Reversible drug resistance and fractional killing in tumor cell population treatment.

We developed a model of drug resistance in TRAIL (TNF-Related Apoptosis Induced-Ligand) treatment in HeLa cell lines. The TRAIL signal transduction pathway is one of the best studied apoptosis pathways and hence permits detailed comparisons with data. Our model was able to explain experimental observations fractional killing and cell-to-cell variability, and predicted reversible resistance [3]. (Work in close collaboration with G. Batt and S. Stoma from the Inria team LIFEWARE.)

6.2.3. Radiotherapy.

Radiation is still a major treatment in cancer. We explored by extensive computer simulations using an agent-based model the consequences of spatially inhomogeneous irradiation. The model predicted that in the case of different competing sub-populations, namely cancer stem cells with unlimited division capacity, and cancer cells with limited division capacity, inhomogeneous radiation focusing higher doses at the tumour centre and lower doses at the tumour periphery should outperform homogeneous irradiation [12]. Cancer stem cells are believed to have a longer cell cycle duration than cancer cells, and are less radiosensitive than cancer cells, which is why they often survive radiation and lead to tumour relapse.

6.2.4. Intercellular interactions in epithelio-mesenchymal transition (EMT).

A PhD thesis on this subject, co-supervised by L. Almeida and M. Sabbah (INSERM team "Cancer biology and therapeutics", St Antoine) has begun at Fall. It is also based on phenotype-structured modelling of Darwinian evolution in cancer cell populations.

6.2.5. Interactions between tumour cell populations and their cellular micro-environment.

A phenotype-structured model of the interactions between a brest cancer cell population (MCF7 cultured cells, collaboration with M. Sabbah, St Antoine Hospital) and its adipocyte stroma support cell population has been developed (T. Lorenzi, J. Clairambault), which, beyond submitted proposals (ANR, Emergence Paris-Sorbonne Universités call), will be studied and experimentally identified in a forthcoming internship (January-June 2015) and PhD thesis in applied mathematics.

---

[0] Sharma *et al.*, Cell, April 2010

6.2.6. Hele-Shaw model of tumour growth.

In the growing field of mathematical analysis of mechanical domain of tumor growth, we focus on the rigorous link between cells models, relying on mechanical properties of cells, and free boundary problem, where the tumor is described by the dynamics of its boundary. The latter model is referred to Hele-Shaw model [44]. Benoît Perthame, Min Tang and Nicolas Vauchelet have proved the rigorous derivation of a geometric model of the Hele-Shaw type for a model with viscoelastic forces, constructing analytically traveling wave solutions of the Hele-Shaw model of tumor growth with nutrient that explain theoretically the numerical results observed. The limiting model exhibits travelling waves, which have been investigated in [43]. Another interesting feature for this model is the transversal instability occurring when the spatial dimension is greater than 1. Together with Fernando Quirós (Univ. Autónoma de Madrid), the aforementioned have also formulated a Hele-Shaw type free-boundary problem for a tumor growing under the combined effects of pressure forces, cell multiplication and active motion, the latter being the novelty of this study [61]. In order to understand the emergence of instabilities in the Hele-Shaw model with nutrients, Michal Kowalczyk (Univ. Chile, Santiago), Benoît Perthame and Nicolas Vauchelet have studied a related model of thermo-reactive diffusion where they can study the spectrum of the linearized system around a traveling wave and in which they can compute the transition to instability in terms of a parameter related to the ratio between heat conduction and molecular diffusion. However, the rigorous study of such instabilities for the whole system of equations is not reachable for the moment; only a study for a simplified model has been performed in [39].

6.2.7. Modelling and control of acute myeloblastic leukæmia (AML).

The collaboration with the Disco project-team has been continued, leading to one book chapter [25], four conference proceedings [21], [22], [23], [24] and JL Avila Alonso's PhD thesis defence.

In more detail:
Starting initially from a PDE model of hematopoiesis designed by Adimy *et al.*[0], we have derived several models of healthy or cancer cell dynamics in hematopoiesis and performed several stability analyses.

We have proposed in [25] a new mathematical model of the cell dynamics in acute myeloid leukæmia (AML) which takes into account the four different phases of the proliferating compartment as well the fast self-renewal phenomenon frequently observed in AML. As was the case in [25] this model is transformed into a distributed delay system and was analyzed here with input-output techniques. Local stability conditions for an equilibrium point of interest are derived in terms of a set of inequalities involving the parameters of the mathematical model.

We have also studied a coupled delay model for healthy and cancer cell dynamics in AML consisting of two stages of maturation for cancer cells and three stages of maturation for healthy cells. For a particular healthy equilibrium point, locally stability conditions involving the parameters of the mathematical model have been obtained [22], [23].

We have performed in [21] a stability analysis of both the PDE model of healthy hæmatopoiesis and a coupled PDE model of healthy and cancer cell dynamics. The stability conditions obtained here in the time domain strengthen the idea that fast self-renewal plays an important role in AML.

A time-domain stability analysis by means of Lyapunov-Krasovskii functionals has been performed on the delay system modeling healthy hematopoiesis for a strictly positive equilibrium point of interest.

Furthermore, a working collaboration on AML modelling with Anna Marciniak-Czochra (Univ. Heidelberg) was also initiated by the end of 2014 by a visit of three of us (C. Bonnet, J. Clairambault, T. Lorenzi) to Heidelberg and a visit of T. Stiehl, A. Marciniak-Czochra PhD student, to Paris. The topics we plan to investigate are, beyond the role of fast self renewal in AML cell populations, the part played by clonal heterogeneity in leukæmic cell populations and the issues it raises in therapeutics, a well known clinical problem in clinical hæmatology.

---

[0] Adimy, M., Crauste, F., El Abllaoui, A. Discrete maturity-structured model of cell differentiation with applications to acute myelogenous leukemia, *J. Biol. Sys.*, 16(3):395-424, 2008

Let us also mention that on the subject of early leukæmogenesis, Andrada Qillas Maran has undertaken a PhD thesis under the supervision of J. Clairambault and B. Perthame. Models relying on piecewise deterministic Markov processes (PDMPs), designed and studied by R. Yvinec (INRA Tours) for the single-cell part of the model under construction, will be used in collaboration with him. Our clinical referents in hæmatology for this PhD work are F. Delhommeau and P. Hirsch (St Antoine Hospital).

6.2.8. The p53 protein spatio-temporal dynamics.

The development of our molecular-based model of the spatio-temporal intracellular dynamics of the p53 protein (the so-called “guardian of the genome") has been continued [55], [9], leading us also, more generally, to propose a modelling frame dedicated to the dynamics of intracellular proteins and their gene regulatory networks [8].

6.2.9. Others.

In a collaboration with ANGE, B. Perthame has studied a data assimilation algorithm for multidimensional hyperbolic conservation laws using kinetic schemes and kinetic formulations.

6.3. Aggregation kinetics

Participants: Tom Banks, Thibault Bourgeron, Marc Hoffmann, Marie Doumic-Jauffret, Nathalie Krell, Benoît Perthame, Stéphanie Prigent, Human Rezaei, Nathalie Robert, Léon Matar Tine [Univ. Lyon and Dracula Inria team], Jorge Zubelli [IMPA, Rio de Janeiro].

6.3.1. Time Asymptotics for Fragmentation Equations

Fragmentation and growth-fragmentation equations is a family of problems with varied and wide applications. This paper is devoted to description of the long time time asymptotics of two critical cases of these equations, when the division rate is constant and the growth rate is linear or zero. The study of these cases may be reduced to the study of the following fragmentation equation:

$$\frac{\partial}{\partial t}u(t,x)+u(t,x)=\int_x^\infty k_0(xy)u(t,y)dy.$$

Using the Mellin transform of the equation, we determine the long time behavior of the solutions. Our results show in particular the strong dependence of this asymptotic behavior with respect to the initial data.

6.3.2. Estimating the division rate in a size-structured population.

The problem which was considered in [5] consists in estimating the division rate from the stable size distribution of the population, which is easily measured, but non-smooth. We propose a method based on the Mellin transform for growth-fragmentation equations with self-similar kernels. We build a sequence of functions which converges to the density of the population in division, simultaneously in several weighted L^2 spaces, as the measurement error goes to 0. This improves previous results for self-similar kernels [0] and allows us to understand the partial results for general fragmentation kernels [0]. Numerical simulations confirm the theoretical results. Moreover, our numerical method is tested on real biological data, arising from a bacteria growth and fission experiment.

6.3.3. What governs bacterial growth? The “sizer” vs the “timer” model

We applied the previously seen inverse problem methodology [5] to a fundamental biological problem: what governs the bacterial growth?

---

[0] Perthame and Zubelli, *Inv. Prob.*, 2007

[0] Doumic and Tine, *J. Math. Biol.*, 2012

Many organisms coordinate cell growth and division through size control mechanisms: cells must reach a critical size to trigger a cell cycle event. Bacterial division is often assumed to be controlled in this way, but experimental evidence to support this assumption is still lacking. Theoretical arguments show that size control is required to maintain size homeostasis in the case of exponential growth of individual cells. Nevertheless, if the growth law deviates slightly from exponential for very small cells, homeostasis can be maintained with a simple 'timer' triggering division. Therefore, deciding whether division control in bacteria relies on a 'timer' or 'sizer' mechanism requires quantitative comparisons between models and data.

The timer and sizer hypotheses find a natural expression in models based on partial differential equations. Here we test these models with recent data on single-cell growth of Escherichia coli. We demonstrate that a size-independent timer mechanism for division control, though theoretically possible, is quantitatively incompatible with the data and extremely sensitive to slight variations in the growth law. In contrast, a sizer model is robust and fits the data well. In addition, we tested the effect of variability in individual growth rates and noise in septum positioning and found that size control is robust to this phenotypic noise.

Confrontations between cell cycle models and data usually suffer from a lack of high-quality data and suitable statistical estimation techniques. In the study [16] we had overcome these limitations by using high precision measurements of tens of thousands of single bacterial cells combined with recent statistical inference methods to estimate the division rate within the models. We therefore provided the first precise quantitative assessment of different cell cycle models.

6.3.4. Size distribution of amyloid fibrils. Mathematical models and experimental data.

More than twenty types of proteins can adopt misfolded conformations, which can co-aggregate into amyloid fibrils, and are related to pathologies such as Alzheimer's disease. In [15], we surveyed mathematical models for aggregation chain reactions, and discussed the ability to use them to understand amyloid distributions. Numerous reactions have been proposed to play a role in their aggregation kinetics, though the relative importance of each reaction in vivo is unclear: these include activation steps, with nucleation compared to initiation, disaggregation steps, with depolymerization compared to fragmentation, and additional processes such as filament coalescence or secondary nucleation. We have statistically analysed the shape of the size distribution of prion fibrils, with the specific example of truncated data due to the experimental technique (electron microscopy). A model of polymerization and depolymerization succeeds in explaining this distribution. It is a very plausible scheme though, as evidenced in the review of other mathematical models, other types of reactions could also give rise to the same type of distributions.

To clarify how these fibrils are able to incorporate additional units, prion fibril aggregation and disaggregation kinetics were experimentally studied using Static Light Scattering (SLS) [45]. Values that are functions of $\sum i^2 c_i$ (for $i > 0$) with c_i being the concentration of fibrils of size i, were then measured as a function of time. An initial model, adapted from the Becker-Döring system that considers all fibrils to react similarly is not able to reproduce the observed in vitro behaviour. Our second model involves an additional compartment of fibrils unable to incorporate more prion units. This model leads to kinetic coefficients which are biologically plausible and correctly simulates the first experimental steps for prion aggregation.

In the formation of large clusters out of small particles, the initializing step is called the nucleation, and consists in the spontaneous reaction of agents which aggregate into small and stable polymers called nucleus. After this early step, the polymers are involved into a bunch of reactions such as polymerization, fragmentation and coalescence. Since there may be several orders of magnitude between the size of a particle and the size of an aggregate, building efficient numerical schemes to capture accurately the kinetics of the reaction is a delicate step of key importance. In [29], we propose a conservative scheme, based on finite volume methods on an adaptive grid, which is able to render out the early steps of the reaction as well as the later chain reactions.

6.4. Liver organ modelling

Participants: Noémie Boissier, Dirk Drasdo, Géraldine Cellière, Adrian Friebel, Group Heinzle [Univ. Saarbruecken, Germany], Group Hengstler [IfADo, Germany], Stefan Hoehme, Tim Johann, Group Klingmueller [German Cancer Center, Heidelberg], Johannes Neitsch, Group Reo [Inria Paris - Rocquencourt], Paul Van Liedekerke, Eric Vibert [Hopital Paul Brousse], Yi Yin, Group Zerial [Max-Planck Inst. for Molecular Genetics, Dresden, Germany], Groups Iflow, Notox, Vln.

6.4.1. Ammonia detoxification after drug-induced damage.

The model for ammonia detoxification after drug-induced damage (see above) identified a systematic deviation between data and results that would be expected from the current standard model for ammonia detoxification in healthy liver [0] ([17], [6]) (see also comments/editorials in [0]). The findings triggered a series of new experiments identifying reversibility of the glutamate-dehydrogenase reaction in hepatocytes, and in blood (Ghallab et. al., subm.). It could be shown in an animal model that the newly recognized reactions can be therapeutically used to significantly reduce the concentration of toxic ammonia after drug-induced damage. (Work in close collaboration with partners of the project VLN (BMBF, Germany) and EU-NOTOX.

6.4.2. Systematic analysis strategies permitting quantitative conclusions in systems medicine and biology.

Based on the examples from liver regeneration after drug-induced damage [57] [17]) systematic iterative strategies can be inferred to enable identification of mechanisms underlying complex processes in spatial temporal tissue organisation and organ functioning. These use an iterative application of a pipeline of imaging, image analysis and modeling, quantitative models by parameterization of model components by measurable parameters for which the physiological ranges are known, and systematic simulated parameter sensitivity analyses [7].

---

[0] Haeussinger D., Eur. J. Biochem, 1983; Gebhardt R and Mecke, D. EMBO J 1983

[0] Wierling, C. Hepatology, 60(6) 2014; and: Widera, A., EXCLI Journal, 13, 2014

MASAIE Project-Team

5. New Results

5.1. Highlights of the Year

The estimation of sequestered parasite population has been a challenge for the biologist and modeler, with many authors having studied this problem. The difficulty is that the infected erythrocyte leaves the circulating peripheral blood and binds to the endothelium in the microvasculature of various organs. A measurement of Plasmodium falciparum parasitaemia taken from a blood smear therefore samples young parasites only and there is no clinical methods to measure the sequestered parasites. We have developed a simple tool to estimate the sequestered parasites and hence the total parasite burden for *Plasmodium falciparum* malaria patients. We have also given a method to estimate a crucial parameter in the model of infection. This parameter β can be thought as the “transmission/invading" factor between merozoites and erythrocytes. This work [9] has been published in "Mathematical Biosciences and Engineering".

5.2. Modeling the use of Wolbachia for controlling the incidence of dengue

We continued research on modeling the introduction of *Wolbachia* in a population of *Aedes Aegypti*. This research is done in collaboration with FGV (Fundação Getulio Vargas), Fiocruz (Fondation Oswaldo Cruz) and UFF (Universidade Federal Fluminense) in Rio de Janeiro (Brazil) [16].

Wolbachia is a bacteria which infects arthropod species, including a high proportion of insects (60% of species). Its interactions with its hosts are often complex, and in some cases it is considered as an endosymbiont. The unique biology of *Wolbachia* has attracted a growing number of researchers interested in questions ranging from the evolutionary implications of infection through to the use of this agent for pest and disease control: a public web site has been funded by the National Science Foundation of Australia, and a research in pubmeb (http://www.ncbi.nlm.nih.gov/pubmed) typing `wolbachia` gives 1889 results.

While *Wolbachia* is commonly found in many mosquitoes it is absent from the species that are considered to be of major importance for the transmission of human pathogens. The successful introduction of a life-shortening strain of *Wolbachia* into the dengue vector *Aedes aegypti* that decreases adult mean life has recently been reported.

Moreover it is estimated that the population of mosquitoes harboring *Wolbachia* is less efficient to transmit dengue [18], [21], [22], [25]. Then it is considered that using *Wolbachia* can be a viable option for controlling the incidence of the dengue.

We consider an alternative infection (by Wolbachia) model which exhibits monotonous properties. This model is designed to take into account both the biology of this infection and any available data. The objective is to use this model for predicting the sustainable introduction of this bacteria. We provide a complete mathematical analysis of the model proposed and give the basic reproduction ratio $\mathcal{R}_0$ for *Wolbachia*. We observe a bistability phenomenon. Two equilibria are asymptotically stable: an equilibrium where all the population is uninfected and an equilibrium where all the population is infected. A third unstable equilibrium exists. We provide a lower bound for the basin of attraction of the desired infected equilibrium. We are in a backward bifurcation situation. The bistable situation occurs with natural biological values for the parameters. Our model is an example of an epidemiological model with only vertical transmission.

This infection model is then connected with a classical dengue model. We prove that for the complete model the equilibrium with *Wolbachia* for the mosquitoes and without dengue for the human is asymptotically stable. We prove that, if a sufficiently large population of infected (by Wolbachia) mosquitoes is introduced, dengue will disappear.

These results have been obtained in collaboration with Pierre-Alexandre Bliman (FGV, Inria); Moacyr Silva (FGV), Claudia Codeço (Fiocruz), Max Souza (UFF) and Jair Koiller (FGV).

5.3. Estimating the proportion of susceptible individuals for a dengue epidemic

Starting from the multi-scaled dengue system, we construct a pair of observers to estimate the dynamics of the disease. The nature of both the observers and the multi-scaled system allows to estimate both the number of susceptible and recovered hosts, as well as to provide information on the vector population, using only infected population data. Numerical simulations have been used to illustrate the performance of the observers.

5.4. Singular value decomposition in dynamic epidemiology: arboviral diseases with human circulation

We introduce a matrix that combines information about human circulation and the epidemiological situation at the nodes of a metapopulational model for an arboviral disease. Its singular value decomposition allows relationships between three basic reproduction numbers $\mathcal{R}_0$: local(s), uniform, and network. The onset of an arboviral disease is strongly dependent on the network characteristics. We present a naive "early warning" criterion for the outbreak at a given node, aiming to promote a discussion on the role of left and right singular vectors. This work is done by the Brazil-France Capes/Cofecub team.

5.5. Analysis of a schistosomiasis infection model

The global mathematical analysis of a schistosomiasis infection model that involves human and intermediate snail hosts as well as an additional mammalian host and a competitor snail species has been done by constructing Lyapunov functions and using properties of K monotone systems. We derived the basic reproduction number $\mathcal{R}_0$ for the deterministic model, and establish that the global dynamics are completely determined by the values of $\mathcal{R}_0$. This mathematical analysis of the model gives insight about the epidemiological consequences of the introduction of a competitor resistant snail species. We gave the characteristics of the competitor resistant snail species that can be used to eliminate the disease [11].

5.6. Multi-stages and multi-strains epidemic models

The model SI (Susceptible-Infected) is one of the most important and used epidemiological models. We gave a complete analysis of the stability of the model with a non-linear incidence and two classes of infected individuals [12].

We have also studied SIS, SIR and MSIR models with bilinear incidence and varying population, with n different pathogen strains of an infectious disease, with or without vertical transmission. For these classes of models, we have proved that under generic conditions a competitive exclusion principle holds. To each strain a basic reproduction ratio can be associated. It corresponds to the case where only this strain exists. The basic reproduction ratio of the complete system is the maximum of the individual basic reproduction ratios. Actually we have also defined an equivalent threshold for each strain. The winner of the competition is the strain with the maximum threshold. It turns out that this strain is the most virulent, i.e., this is the strain for which the endemic equilibrium gives the minimum population for the susceptible host population. This can be interpreted as a pessimization principle [10].

A mathematical multi-patches model for highland malaria in Kenya has been developed and analysed in [13] and [14].

MODEMIC Project-Team

6. New Results

6.1. Highlights of the Year

Yeasts play a central role in the wine making process. To study the yeasts in a stable environment and physiological state, a Multi-Stage Continuous Fermentor (MSCF) has been designed by the research Unit SPO (Sciences For Oenology). This device mimics the steps of the batch fermentation process. In this paper, the problem of the control of the sugar concentrations in each of the four reactors of the MSCF is considered. The cascade structure of the device leads to a constraint on the input flow rates (the control variables). A control strategy based on a linearizing control law coupled with a state observer and an anti windup component is proposed and finally implemented on the experimental process (see also 6.3.2).
BEST PAPER AWARD :
[41] **19th IFAC World Congress 2014**. C. CASENAVE, D. DOCHAIN, J. HARMAND, M. PEREZ, A. RAPAPORT, J.-M. SABLAYROLLES.

6.2. Mathematical models for microbial ecology

6.2.1. Differential equations models

Participants: Céline Casenave, Jérôme Harmand, Claude Lobry, Alain Rapaport, Alejandro Maximiliano Rojas.

Anaerobic digestion refers to the transformation of biodegradable material by micro-organisms in absence of oxygen (it can be found in waste-water treatments or industrial fermentation, and occurs naturally in soils). It receives an increasing consideration due to recent technological advances, but also because it is a source of renewable energy (bio-gas, fuel...). The anaerobic digestion is a complex set of bio-processes, for which there is a strong expectation of tractable models. We have proposed and studied new mathematical models that takes into account the following features:

- The available anaerobic digestion models used for control purposes do usually only consider soluble matter. In fact, part of the pollutants are not soluble but are under a particulate form. In order to establish whether adding the dynamics of such matter into the models is important for the system behavior or not, we have studied new models and established that depending on the kinetics of this additional reaction step, the qualitative behavior of the process may be significantly modified [26].
- Microbial food chains are present in anaerobic digestion where the different reaction steps can be seen as such: the waste products of the organisms at one trophic level (i.e. one reaction step) are consumed by organisms at the next trophic level (i.e. the next reaction step). In [55] we study a model of a two-tiered microbial 'food chain' with feedback inhibition, which was recently presented [63] as a reduced and simplified version of the anaerobic digestion model ADM1 of the International Water Association (IWA). It is known that in the absence of maintenance (or decay) the microbial 'food chain' is stable. In [63], using a purely numerical approach and ADM1 consensus parameter values, it was shown that the model remains stable when decay terms are added. In [55] we prove that introducing decay in the model preserves stability whatever its parameters values are and for a wide range of kinetics.

For the study of spatial heterogeneity in the models, we have carried on mathematical analyses of the properties of interconnected chemostats, in particular when growth rates present a substrate inhibition. In addition to the stabilizability properties discovered last year on "buffered" interconnections [33], we have studied this year yielding performances at steady state for the stabilizing configurations and characterized the set of the most efficient ones. For such configurations, we have shown that under certain circumstances, a "by-pass" of the main tank could be the best solution to ensure a global stabilization.

We have also analyzed two kinds of models, suited to specific characteristics of the microbial activity in soils:

- In [30], we have studied analytically and numerically a piece-wise linear model of carbon mineralization by two functional groups of micro-organisms in view of predicting the "priming effect" in soil ecosystems. The conclusion is that under a climate change, the augmentation of $C0^2$ will not affect to primary production and carbon storage when the plants are limited by nitrogen, but surprisingly a higher carbon input in soil should lead to a deplete of sequestered carbon and the increase of nitrogen release.
- In collaboration with Géosciences Rennes (Jean-Raynald de Dreuzy, Tristan Babey) and in the scope of the co-supervision of the PhD of Alejandro Rojas (also in the collaboration within the associated team with Chile), we investigate the equivalence between networks that represent interconnections of mobile/immobile zones in mass transfer models for soil ecosystems. For Structured INteracting Continua (SINC) models, that are described as the combination of a finite number of diffusion-dominated interconnected immobile zones exchanging with an advection-dominated mobile domain, we have proved an equivalence with Multi-Rate Mass Transfer (MRMT) and proposed a method for the identification of the equivalent MRMT model [14]. Moreover, we have shown the role of the controllability properties of a sub-system, in addition to the irreducibility of the network graph, for the input/output equivalence between several representations (work in preparation).

6.2.2. Stochastic and hybrid discrete-continuous dynamical models

Participants: Fabien Campillo, Bertrand Cloez, Coralie Fritsch.

Hybrid mass-structured chemostat models

Within the context of Coralie Fritsch thesis [12], we adopt a new modeling approach where instead on focusing on one type of model we propose different models and their interconnections, on the numerical viewpoint as well as the analytical one. Namely we propose an hybrid model of the chemostat where the population of bacteria is individually-based, each individual being described by its mass, and the subtract concentration is represented as a classic differential equation. We proved the convergence of this model in high population size toward an integro-differential system [20]. We proposed specific numerical schemes for the two approaches (see 5.3) [27].

Evolutionary invasion analysis and simulation for the chemostat

Still in the context of Coralie Fritsch thesis [12] and following her last year stay at the University of Helsinki in Otso Ovaskainen's Research Group, we consider an hybrid mass-structured mass-structured chemostat models with trait. The trait could for example be the factor of mass dissymmetry in the binary fission of a bacteria. In this context we proved an equivalence between invasion fitness for the hybrid-IBM model and the integro-differential system. We also numerically exhibit an evolutionarily singular strategy: with this given trait a monomorphic resident population cannot be invaded by a mutated population; the result is true for the two models.

PDE and stochastic models

In collaboration with M. Joannides and I. Larramendy-Valverde (I3M, University of Montpellier) we consider a stochastic growth model for which extinction eventually occurs almost surely. The associated complete Fokker–Planck equation describing the law of the process is established and studied. In dimension one, e.g. for the stochastic logistic model this equation combines a PDE and an ODE (paper under revision); in dimension two, e.g. for the stochastic chemostat model this equation combines a 2D PDE and a 1D PDE [22]. We then design a finite differences numerical scheme under a probabilistic viewpoint.

6.2.3. Other modeling approaches

Participants: Anne Bisson, Jérôme Harmand, Alain Rapaport.

A collaboration with the UMR Eco & Sols has led to the development of a (static) probabilistic model for inferring nature and number of interactions in communities assembly [29]. This model has brought new insights on a data set from reconstituted soil ecosystems. Because of the curse of dimensionality, we have begun this year to extend this approach to "assembling motifs" instead of considering all the possible assemblages (paper in preparation).

In [28], ecological trade-offs between species are studied to explain species coexistence in ecological communities. In our model, plant species compete for sites where each site has a fixed stress condition. Species differ both in stress tolerance and competitive ability. We derive the deterministic discrete-time dynamical system for the species abundances. We prove the conditions under which plant species can coexist in an stable equilibrium. We compare our model with a recently proposed, continuous-time dynamical system for a tolerance-fecundity trade-off in plant communities, and we show that this model is a special case of the continuous-time version of our model.

6.3. Analysis and supervision of bioprocesses

6.3.1. Models development and identification

Participants: Fabien Campillo, Amine Charfi, Yessmine Daoud, Jérôme Harmand, Sonia Hassam, Guilherme Pimentel, Alain Rapaport.

Membrane bioreactors combine a filtration process (with a membrane) and a suspended growth rate bioreactor. This recent technology present many advantages compared to conventional ones, but is more sophisticated and requires refined control because of the fouling process. We have proposed new modeling approaches of such bioreactors, where a fouling mechanism is explicitly described:

- Membrane bioreactors allow a perfect separation of biomass and treated waters. However, membranes are subject to clogging - also called fouling - by large organic molecules and solids (biomass, suspended solids, etc...). This phenomenon represents the main problem that limits the development of membranes bioreactors. It is well documented but very studies tried to formalize it under the form of dynamical model, even more under anaerobic conditions. This is what we did in the framework of Amine Charfi's PhD [11]: he contributed to a better understanding of membrane fouling dynamics and proposed a number of strategies to avoid irreversible clogging [24], [25].
- In the scope of the PhD supervision of Guilherme Pimentel, we have proposed a simple three time scales model in view of the control of the cake formation [13], [37]. This model has been validated on real data from a pilot plant at Univ. Mons (Belgium).

As already mentioned in 6.2.1 , anaerobic bioreactors are able to produce valuable energy. However, they are subject to destabilization in case of organic overload. It is thus necessary to develop appropriate models dedicated to the synthesis of stabilizing control feedbacks. Two strategies are followed to obtain such models. Either simple models are proposed from the knowledge we have about the process, either reduced models are obtained from more complex ones. Yessmine Daoud just began her PhD thesis following the first strategy: more precisely, we study simple inhibition models of the anaerobic digestion and tries to establish simple equivalence between these models and the well known ADM1 model. Sonia Hassam, in her PhD, works within the framework of the second strategy to propose simple models obtained in reducing complex ones like the ADM1.

As already proposed last year, we have set a methodology to identify from data observed on a chemostat plant a multi-specific model that suits better than a mono-specific one, when data from molecular biology are available. In [35], we combine molecular fingerprints obtained at some discrete times (such as the ones provides by the DGGE or SSCP techniques) with on-line macroscopic measurements. In a similar spirit, within the framework of a collaboration with the LOMIC at Banyuls, we have analysed molecular data obtained via SSCP technique to monitor the structure of microbial communities. It was shown that aerial transport of bacteria from desert into the sea influenced its bacterial diversity [32], [31].

As far as stochastic approaches are concerned, the thesis of Mohsen Chebbi aims at developing stochastic models of membrane bioreactors following the approach proposed in [57]. A mathematical and simulation framework has been established, as well as the bases of vectorial simulation techniques in `Matlab`. Developments of Monte Carlo techniques for the identification of bioprocesses are investigated in the thesis of Oussama Hadj-Abdelkader, improving classical particle filtering approaches (sequential Monte Carlo) by integrating MCMC (Monte Carlo Markov Chain) procedures. A software approach has been adopted in `C++`.

6.3.2. Synthesis of control laws

Participants: Térence Bayen, Walid Bouhafs, Céline Casenave, Amel Ghouali, Jérôme Harmand, Zeyneb Khedim, Claude Lobry, Alain Rapaport, Victor Riqueleme, Matthieu Sebbah.

We investigate two kinds of bioprocesses to be controlled, arising in industrial biotechnology (digesters, waste-water purification...) or in the bioremediation of natural environments (lakes, landfill...).

6.3.2.1. Industrial biotechnology

As it is often the case in industry, we distinguish two kinds of process operating: continuous processes, for which the volumes of the bio-reactors are constant, and fed-batch processes, for which the filling rate is the control.

We tackle several optimal control problems related to the maximization of productivity of continuous bioprocesses:

- As far as anaerobic digestion is concerned, we have considered an optimal control problem for the chemostat model with substrate inhibition. The originality of this problem relies on the fact that the quantity to be maximized is not simply a state of the model (the substrate or the biomass) but the volume of biogas produced (which is a nonlinear function of the state) within a given time interval. Optimal strategies have been proposed for a class of initial conditions of the system (PhD thesis of Amel Ghouali). Other optimal control problems are studies by Walid Bouhafs to establish the optimality of controls initially proposed by Djalel Mazouni, a former PhD candidate supervised by the team. Using a tricky projection of the problem into another state space, he has shown that the problem of optimizing the degradation of two different substrates by two antagonist bacterial communities could be solved in a very general sense as the minimization of a specific functional for a very large class of nonlinear systems [19].
- The work [15] arises in the context of selection of species (widely used in agriculture and biotechnology in order to improve productivity). For microorganisms, the selection process can be based on genetic tools. Our methodology in [15] is to drive the competition between species in a chemostat. We consider a two species chemostat model with one limiting substrate, and our aim is to optimize the selection of the species of interest. Thanks to the Pontryagin Maximum Principle, we introduce a singular feeding strategy which allows to reach the target, and we prove that the feedback control provided by this strategy is optimal. The optimal synthesis of the problem in presence of more than two species will be investigated in a future work.
- In [38], we study the problem of minimal time for a chemostat system with one limiting substrate and one species. Given a target point, the problem consists in finding an optimal feeding strategy steering any initial condition of the system to this target. This is typically of interest whenever the input substrate concentration changes yielding in a new steady state. We consider the case where the growth rate function is of Haldane type implying the existence of a singular arc that is non-necessary admissible everywhere. We provide an optimal synthesis of the problem using tools from optimal control theory.
- The work [16] studies the coupling of a culture of micro-algae limited by light and an anaerobic digester in a two-tank bioreactor (the model combines a periodic day-night light for the culture of micro-algae and a classical chemostat model for the digester). We first prove the existence and attraction of periodic solutions of this problem for a one day period. Then, we study the optimal control problem of optimizing the production of methane in the digester during a certain time frame, the control on the system being the dilution rate (the input flow of micro-algae in the digester). We also investigate the dependence of the optimal cost with respect to the volume ratio of the two tanks.

In collaboration with researchers of the unit SPO (Sciences For Oenology), we have proposed a control law of a multi-stage continuous fermentor (MSCF) designed for the study of the wine fermentation, that has been implemented on the experimental process [41]. We have also finalized the controllability analysis and minimal time feedback synthesis of models of cascade of continuous bioreactors under input constraints [18], [40] (that is also related to the control of MSCF).

As far as fed-batch processes are concerned, an extension of former results of the team about the minimal time control of fed-batch processes with impulse controls is presented in [34],

The paper [17] is devoted to the study of the minimal time problem of a fed-batch reactor, under the presence of a saturation point on the singular locus (this typically occurs whenever the growth rate function is of Haldane type and when typically the maximum input flow rate is not high enough to maintain the substrate concentration constant). This brings non-intuitive issues for the optimal synthesis (existence of switching curve and point of prior saturation).

6.3.2.2. Bioremediation of natural environments

In the scope of the associated team with Chile and the supervision of the postdoctoral stay of Matthieu Sebbah in Chile, we have addressed a new model of landfill remediation when controlling the leachate recirculation [53]. We have applied the same methodological approach than the one for the work [18], [40] mentioned in the previous section, which consists in characterizing first the sub-domains for which the target can be optimally reached with a constant extreme control (no recirculation or maximal speed of recirculation), and further the nature of optimal commutations outside these sets. This analysis provides information for the practitioners on the benefit to implement sensors and real-time controllers.

Also in the scope of the associated team with Chile (see 7.3.2.1) and the co-supervision of the PhD of Victor Riquelme, we have carried on the study of optimal syntheses for the minimal time treatment of natural water reservoirs (such as lakes) [52]. We have proved that the minimal time strategy consists in a most-rapid approach to homogeneous concentrations, even though the optimal control problem is non convex. Moreover, we have shown that a large diffusion increases the treatment time when the resource is everywhere highly polluted , while it can at the opposite be beneficial when only part of the resource is polluted. This feature should serve the practitioners in the choice of pumps positioning in a originally clean water resource that is suddenly affected by a local pollution. This work is in connection with the INRA/Inria patent [47] that has been deposited jointly with LEMON Team.

6.4. Other application domains

Participants: Fabien Campillo, Céline Casenave.

Semi-Markov land use dynamic

With IRD (GRED Montpellier) and the Univ. of Fianarantsoa (Madagascar) we pursued our study on land use dynamics models corresponding to parcels located on the edge of the forest corridor, Madagascar. We use semi-Markov chain to infer the land-use dynamics. In addition to the empirical and maximum likelihood methods, we estimate the semi-Markov kernel by a Bayesian approach [21].

Ice cream crystallization

We study the problem of the control of an ice cream crystallization process, part of the European CAFE project, in collaboration with CESAME (Univ. Catholique de Louvain-la-neuve), Irstea Antony and AgroParisTech. The goal is to control the viscosity of the ice cream at the outlet of the continuous crystallizer. On the basis of a population balance equation describing the evolution of the crystal size distribution of the ice cream, and an energy balance equation, we have proposed an input-output reduced order model of the process, that has been identified and validated on experimental data [23]. A nonlinear control strategy based on an adaptive linearizing control law coupled with a Smith predictor to account for the measurement delay has been proposed, and validated on the experimental pilot plant (paper in preparation).

MYCENAE Project-Team

6. New Results

6.1. Highlights of the Year

- Picture of the Conference poster of the 2014 SIAM annual meeting (July 7-11, Chicago, USA), adapted from [7]
- Invitation to organize the mini symposium “The stochastic brain” at the Stochastic Processes and Applications Conference (Jul 28-Aug1, Buenos-Aires, Argentina)
- Selection of the NeuroMathMod project in the framework of the Sorbonne Université Emergence 2014 call

6.2. Numerical and theoretical studies of slow-fast systems with complex oscillations

6.2.1. A multiple time scale coupling of piecewise linear oscillators: Application to a neuroendocrine system

Participants: Frédérique Clément, Mathieu Desroches, Soledad Fernández García, Maciej Krupa.

We have analyzed a four dimensional slow-fast piecewise linear system consisting of two coupled oscillators [32]. Each oscillator is a continuous slow-fast piecewise linear system with three zones of linearity. The coupling is one-way, that is, one subsystem evolves independently and is forcing the other subsystem. We have analyzed not only the qualitative behavior, but also quantitative aspects such as the period, frequency and amplitude of the oscillations. The system is used to reproduce all the features endowed in a former smooth model and reproduce the secretion pattern of the hypothalamic neurohormone GnRH along the ovarian cycle in different species.

6.2.2. Border collision bifurcations of stroboscopic maps in periodically driven spiking models

Participants: Frédérique Clément, Albert Granados Corsellas, Maciej Krupa.

In [21], we have considered a general nonautonomous hybrid system based on the integrate-and-fire model, widely used as simplified version of neuronal models and other types of excitable systems. Our assumptions are that the system is monotonic, possesses an attracting subthreshold equilibrium point, and is forced by means of a periodic pulsatile (square wave) function. In contrast to classical methods, in our approach we use the stroboscopic map (time-T return map) instead of the so-called firing map. It becomes a discontinuous map potentially defined in an infinite number of partitions. By applying theory for piecewise-smooth systems, we avoid relying on particular computations, and we develop a novel approach that can be easily extended to systems with other topologies (expansive dynamics) and higher dimensions. More precisely, we have rigorously studied the bifurcation structure in the two-dimensional parameter space formed by the amplitude of the pulse and the ratio between T and the duration of the pulse (duty cycle). We show that it is covered by regions of existence of periodic orbits given by period adding structures. The period adding structures completely describe not only all the possible spiking asymptotic dynamics but also the behavior of the firing rate, which is a devil’s staircase as a function of the parameters.

6.2.3. Interpreting frequency responses to dose-conserved pulsatile input signals in simple cell signaling motifs

Participants: Richard Bertram, Patrick Fletcher, Joël Tabak [Florida State University], Frédérique Clément, Alexandre Vidal.

Many hormones are released in pulsatile patterns. This pattern can be modified, for instance by changing pulse frequency, to encode relevant physiological information. Often other properties of the pulse pattern will also change with frequency. How do signaling pathways of cells targeted by these hormones respond to different input patterns? We have asked if a given dose of hormone can induce different outputs from the target system, depending on how this dose is distributed in time [20]. We have used simple mathematical models of feedforward signaling motifs to understand how the properties of the target system give rise to preferences in input pulse pattern. We frame these problems in terms of frequency responses to pulsatile inputs, where the amplitude or duration of the pulses is varied along with frequency to conserve input dose. We have found that nonlinearity in the steady state input-output function of the system predicts the optimal input pattern. It does so by selecting an optimal input signal amplitude. Our results predict the behavior of common signaling motifs such as receptor binding with dimerization, and protein phosphorylation. The findings have implications for experiments aimed at studying the frequency response to pulsatile inputs, as well as for understanding how pulsatile patterns drive biological responses via feedforward signaling pathways.

6.2.4. *Mixed-mode oscillations due to a singular Hopf bifurcation in a forest pest model*

Participants: Morten Brøns [Technical University of Denmark], Mathieu Desroches, Maciej Krupa.

We have revisited a three-dimensional model of forest pest where MMOs play an important role [17]. In this model, young trees are distinguished from old trees, and the pest feeds on old trees. The pest grows on a fast scale, the young trees on an intermediate scale, and the old trees on a slow scale. We have established that the main organizing center for the shape and oscillatory patterns of the solutions is not a folded-node singularity, which does exist in the system, but rather a singular Hopf bifurcation. A combination of a singular Hopf bifurcation and a weak return mechanism, characterized by a very small change in one of the variables, determines the features of the mixed-mode oscillations. Period-doubling and saddle-node bifurcations lead to closed families (called isolas) of periodic solutions in a bifurcation corresponding to a singular Hopf bifurcation.

6.2.5. *On the Dynamics of the adenylate energy system: homeorhesis versus homeostasis*

Participants: Jesús M Cortés, Ildefonso M. de La Fuente, Iker Malaina, Luis Martínez, Edelmira Valero [University of Bilbao], Serafim Rodrigues [Plymouth University], Mathieu Desroches.

We have developed and analyzed a new model of the ATP-ADP-AMP biochemical system in order to understand some of the functional elements involved in the cellular energy status [18]. In this model based on a delayed differential system, the enzymatic rate equations and all the physiological kinetic parameters have been explicitly considered and experimentally tested in vitro. Our central hypothesis is that cells are characterized by changing energy dynamics (homeorhesis). The results have shown that the adenylate energy charge (AEC) presents stable transitions between steady states and periodic oscillations and, in agreement with experimental data these oscillations range within the narrow AEC window. Furthermore, the model shows sustained oscillations in the Gibbs free energy and in the total nucleotide pool.

6.2.6. *Adaptative algorithms for the simulation of slow-fast coupled oscillators in networks*

Participants: Frédérique Clément, Marie Postel, Alexandre Vidal.

The numerical simulation of a slow fast system is usually performed using an explicit scheme with an adaptive time step, in order to preserve the numerical accuracy during the fast dynamic events. In the case of large sized networks of coupled slow-fast systems, one need to use the same very small time step for all components of the network, since the integration is performed simultaneously on the whole network. We have proposed a new algorithm based on a dynamic split of the network components, in the framework of symplectic integrators [40], and applied it to a model describing the intracellular calcium oscillations in a network of embryonic GnRH neurons [9]. At each time step, the systems currently in the fast dynamic parts, are identified from their distance to the fast manifold. These components are accordingly integrated using a small time step, while a larger time step is used for the remaining of the network (cf poster abstract in the CANUM 2014 conference). Although the CPU time saving is proportional to the time constant ratio between the slow and fast dynamics, it hardly compensates the drop in the convergence order as the size of the network increases.

6.3. Non conservative transport equations for cell population dynamics

6.3.1. Adaptive mesh refinement strategy for a nonconservative transport problem

Participants: Benjamin Aymard, Frédérique Clément, Marie Postel.

In the framework of transport equations it is usual to need long time simulations, and therefore large physical domains to cover a phenomenon. On the other hand it can happen that only a small time varying portion of the domain is interesting. This motivates the use of adaptivity for the spatial discretization. Biological models involving cell development are often nonconservative to account for cell division. In that case the threshold controlling the spatial adaptivity may have to be time-dependent in order to keep up with the progression of the solution. In [16], we tackle the difficulties arising when applying a Multiresolution method to a transport equation with discontinuous fluxes modeling localized mitosis. The analysis of the numerical method is performed on a simplified model and numerical scheme. An original threshold strategy is proposed and validated thanks to extensive numerical tests. It is then applied to a biological model in both cases of distributed and localized mitosis.

6.3.2. Calibration of a multiscale model for cell dynamics

Participants: Benjamin Aymard, Frédérique Clément, Marie Postel, Kim Long Tran.

In the framework of the PhD of Benjamin Aymard and the master training of Kim Long Tran, we have tackled the issue of the numerical calibration of our multiscale model of cell populations in ovarian follicles, in collaboration with Danielle Monniaux (INRA Tours). The strategy has consisted in designing quantitative specifications from the available biological knowledge, most of which fall within the field of cell population kinetics (e.g. growth fraction, mitotic index ...), and translating them into constraints on the model parameters, as well as in performing a detailed a priori analysis of the properties of the mathematical functions entering the model equations. Using visualization approaches appropriate both for following the trajectory of a given ovarian follicle with time and comparing the follicles together, we have confronted the model outputs on different levels (from the local cell density to the overall cell number) to the corresponding specifications. We have been able to reproduce instances of the selection process occurring within a cohort of terminally growing follicles. To enable one to do systematic explorations of the model behavior in different parameter configurations associated with either physiological (e.g. species-specific ovulation number) or pathological situations (dysovulation), we have undertaken a reduction approach inspired from [41]. We have generalized these results by relaxing some simplifying assumptions to account for some important features of the original model as the distinction between different phases in the cell division cycle.

6.4. Macroscopic limits of stochastic neural networks and neural fields

6.4.1. Pulsatile localized dynamics in delayed neural-field equations in arbitrary dimension

Participants: Jonathan Touboul, Grégory Faye [EHESS].

Neural field equations are integro-differential systems describing the macroscopic activity of spatially extended pieces of cortex. In such cortical assemblies, the propagation of information and the transmission machinery induce communication delays, due to the transport of information (propagation delays) and to the synaptic machinery (constant delays). We have investigated the role of these delays on the formation of structured spatiotemporal patterns for these systems in arbitrary dimensions [19]. We have focused on localized activity, either induced by the presence of a localized stimulus (pulses) or by transitions between two levels of activity (fronts). Linear stability analysis allows to reveal the existence of Hopf bifurcation curves induced by the delays, along different modes that may be symmetric or asymmetric. We show that instabilities strongly depend on the dimension, and in particular may exhibit transversal instabilities along invariant directions. These instabilities yield pulsatile localized activity, and depending on the symmetry of the destabilized modes, either produce spatiotemporal breathing or sloshing patterns.

6.4.2. Limits and dynamics of randomly connected neuronal networks

Participants: Cristóbal Quiñinao [CIRB], Jonathan Touboul.

Networks of the brain are composed of a very large number of neurons connected through a random graph and interacting after random delays that both depend on the anatomical distance between cells. In order to comprehend the role of these random architectures on the dynamics of such networks, we have analyzed the mesoscopic and macroscopic limits of networks with random correlated connectivity weights and delays [35]. We have addressed both averaged and quenched limits, and shown propagation of chaos and convergence to a complex integral McKean-Vlasov equations with distributed delays. We have then instantiated a completely solvable model illustrating the role of such random architectures in the emerging macroscopic activity. We have particularly focused on the role of connectivity levels in the emergence of periodic solutions.

6.4.3. The propagation of chaos in neural fields

Participant: Jonathan Touboul.

We have considered the problem of the limit of bio-inspired spatially extended neuronal networks including an infinite number of neuronal types (space locations), with space-dependent propagation delays modeling neural fields [24]. The propagation of chaos property is proved in this setting under mild assumptions on the neuronal dynamics, valid for most models used in neuroscience, in a mesoscopic limit, the neural-field limit, in which we can resolve the quite fine structure of the neuron activity in space and where averaging effects occur. The mean-field equations obtained are of a new type: they take the form of well-posed infinite-dimensional delayed integro-differential equations with a nonlocal mean-field term and a singular spatio-temporal Brownian motion. We have also shown how these intricate equations can be used in practice to uncover mathematically the precise mesoscopic dynamics of the neural field in a particular model where the mean-field equations exactly reduce to deterministic nonlinear delayed integro-differential equations.

6.4.4. Spatially extended networks with singular multi-scale connectivity patterns

Participant: Jonathan Touboul.

In [24], we took care of a number of technical difficulties arising in the description of large-scale systems that are spatially extended. The organization of neurons in space (within cortical columns) and their interactions (fully connected networks) were relatively far from what is known of the anatomy of neuronal networks. In [25], we have further taken into account the fine and macroscopic structure of the cortex, which is a very large network characterized by a complex connectivity including at least two scales. On the microscopic scale, the interconnections are non-specific and very dense, while macroscopic connectivity patterns connecting different regions of the brain at larger scale are extremely sparse. This motivates to analyze the behavior of networks with multiscale coupling, in which a neuron is connected to its $v(N)$ nearest-neighbors where $v(N) = o(N)$, and in which the probability of macroscopic connection between two neurons vanishes. These are called singular multi-scale connectivity patterns. We have introduced a class of such networks and derived their continuum limit. We show convergence in law and propagation of chaos in the thermodynamic limit. The limit equation obtained is an intricate non-local McKean-Vlasov equation with delays which is universal with respect to the type of micro-circuits and macro-circuits involved.

6.4.5. Index Distribution of the Ginibre Ensemble

Participants: Romain Allez [Stastlab, Cambridge University], Gilles Wainrib [ENS], Jonathan Touboul.

Complex systems, and in particular random neural networks, are often described by randomly interacting dynamical systems with no specific symmetry. In that context, characterizing the number of relevant directions necessitates fine estimates on the Ginibre ensemble. We have computed analytically the probability distribution of the number of eigenvalues N_R with modulus greater than R (the index) of a large $N \times N$ random matrix in the real or complex Ginibre ensemble [15]. We have shown that the fraction $N_R/N = p$ has a distribution scaling as $exp(-\beta N^2 \psi_R(p))$ with $\beta = 1$ (respectively $\beta = 1/2$) for the complex (resp. real) Ginibre ensemble. For any $p \in [0, 1]$, the equilibrium spectral densities as well as the rate function $\psi_R(p)$ are explicitly derived. This function displays a third order phase transition at the critical (minimum) value $p_R^* = 1 - R^2$, associated to a phase transition of the Coulomb gas. We have deduced that, in the central regime, the fluctuations of the index N_R around its typical value $p_R^* N$ scale as $N^{1/3}$.

6.4.6. The heterogeneous gas with singular interaction: Generalized circular law and heterogeneous renormalized energy

Participants: Luis-Carlos Garcia Del Molino, Khashayar Pakdaman [Institut Jacques Monod], Jonathan Touboul.

We have introduced and analyzed d dimensional Coulomb gases with random charge distribution and general external confining potential [23]. Our long term motivation is to understand the spectrum of random matrices with non identical distributions, for instance with independent elements with distinct statistics. We have shown that these gases satisfy a large deviation principle. The analysis of the minima of the rate function (which is the leading term of the energy) reveals that at equilibrium, the particle distribution is a generalized circular law (i.e. with spherical support but non-necessarily uniform distribution). In the classical electrostatic external potential, there are infinitely many minimizers of the rate function. The most likely macroscopic configuration is a disordered distribution in which particles are uniformly distributed (for $d = 2$, the circular law), and charges are independent of the positions of the particles. General charge-dependent confining potentials unfold this degenerate situation: in contrast, the particle density is not uniform, and particles spontaneously organize according to their charge. In that picture the classical electrostatic potential appears as a transition at which order is lost. Sub-leading terms of the energy are derived: we show that these are related to an operator, generalizing the Coulomb renormalized energy, which incorporates the heterogeneous nature of the charges. This heterogeneous renormalized energy informs us about the microscopic arrangements of the particles, which are non-standard, strongly depending on the charges, and include progressive and irregular lattices.

NUMED Project-Team

5. New Results

5.1. Highlights of the Year

Vincent Calvez has been awareded an ERC Grant and the prestigious Bronze medal CNRS.

REO Project-Team

6. New Results

6.1. Highlights of the Year

- Jimmy Mullaert was awarded the best poster prize at the conference Canum 2014.
- Jessica Oakes was awarded a University of California Presidential Postdoctoral Fellowship.
- Jessica Oakes won a young investigator award at the “4th International Conference on Engineering Frontiers in Pediatric and Congenital Heart Disease”.

6.2. Mathematical and numerical analysis of fluid-structure interaction problems

Participants: Benoit Fabrèges, Miguel Ángel Fernández Varela, Mikel Landajuela Larma, Jimmy Mullaert, Marina Vidrascu.

- In [54] we introduce two new classes of numerical methods for the solution of incompressible fluid/thin-walled structure interaction problems with unfitted meshes. The semi-implicit or explicit nature of the splitting in time is dictated by the order in which the spatial and time discretizations are performed. Stability and optimal accuracy are achieved without restrictive CFL conditions or correction iterations. Results presented by M. Landajuela at the 11th World Congress on Computational Mechanics (WCCM XI), July 20-25, 2014, Barcelona (Spain).
- In [47] we introduce a class of fully decoupled time-marching schemes (velocity-pressure-displacement splitting) for the coupling of an incompressible fluid with a thin-walled viscoelastic structure. The time splitting combines a projection method in the fluid with a specific Robin-Neumann treatment of the interface coupling. A priori energy estimates guaranteeing unconditional stability are established for some of the schemes. The accuracy and performance of the methods proposed is illustrated by a thorough numerical study.
- We have performed an a priori error analysis for the generalized Robin-Neumann explicit coupling schemes introduced in [30]. The analysis confirms the $\mathcal{O}(\tau^{2^{r-1}}/h^{\frac{1}{2}})$ error perturbation anticipated by the numerical evidence of [30]. Another fundamental result of this work is that the h-non-uniformity of the splitting error is not a consequence of the mass-lumping approximation (which simply dictates the explicit or semi-implicit nature of the coupling scheme). The analysis indicates that the genesis of the $\mathcal{O}(h^{-\frac{1}{2}})$ is the non-uniformity of discrete viscoelastic operator, which is a consequence of thick-walled character of the solid. These results have been reported in [48] and presented by M.A. Fernández at the 11th World Congress on Computational Mechanics (WCCM XI), July 20-25, 2014, Barcelona (Spain).
- We consider the extension of the Nitsche-XFEM method to fluid-structure interaction problems involving a thin-walled elastic structure (Lagrangian formalism) immersed in an incompressible fluid (Eulerian formalism). The fluid domain is discretized with an unstructured mesh not fitted to the solid mid- surface mesh. Weak and strong discontinuities across the interface are allowed for the velocity and pressure, respectively. The kinematic/kinetic fluid-solid coupling is enforced consistently using a variant of Nitsche’s method involving cut elements. Robustness with respect to arbitrary interface/element intersections is guaranteed through a ghost penalty stabilization. Different coupling schemes, either fully implicit or loosely coupled, are proposed. Several numerical examples, involving static and moving interfaces, illustrate the performance of the methods. A paper in collaboration with F. Alauzet (project-team Gamma3) is under preparation. Results presented by B. Fabrèges at the 11th World Congress on Computational Mechanics (WCCM XI), July 20-25, 2014, Barcelona (Spain).

6.3. Numerical methods for biological flows

Participants: Grégory Arbia, Benoit Fabrèges, Miguel Ángel Fernández Varela, Justine Fouchet-Incaux, Jean-Frédéric Gerbeau, Céline Grandmont, Sanjay Pant, Saverio Smaldone, Marc Thiriet, Irène Vignon-Clementel.

- In [19] We consider the problem of estimating the stiffness of an artery wall using a data assimilation method applied to a 3D fluid-structure interaction (FSI) model. We briefly present the FSI model, the data assimilation procedure based on a reduced order Unscented Kalman filter, and the segmentation algorithm. We then present two examples of the procedure using real data. First, we estimate the stiffness distribution of a silicon rubber tube from image data. Second, we present the estimation of aortic wall stiffness from real clinical data.
- In [29], we propose a new approach to the loosely coupled time-marching of a fluid-fluid interaction problems involving the incompressible Navier-Stokes equations. The methods combine a specific explicit Robin-Robin treatment of the interface coupling with a weakly consistent interface pressure stabilization in time. A priori energy estimates guaranteeing stability of the splitting are obtained for a total pressure formulation of the coupled problem. The performance of the proposed schemes is illustrated on several numerical experiments related to simulation of aortic blood flow.
- In [55] we investigate the stability of numerical schemes that are classically used in the simulation of airflows and blood flows. The geometrical complexity of the networks in which air/blood flows leads to a classical decomposition of two areas: a truncated 3D geometry corresponding to the largest contribution of the domain, and a 0D part connected to the 3D part, modelling air/blood flows in smaller airways/vessels. The resulting Navier-Stokes system in the 3D truncated part may involve non-local boundary conditions, deriving from a mechanical model. For various 3D/0D coupled models, different discretization processes are presented and analyzed in terms of numerical stability, highlighting strong differences according to the regimes that are considered. In particular, two main stability issues are investigated: first the coupling between the 3D and the 0D part for which implicit or explicit strategies are studied and, second, the question of estimating the amount of kinetic energy entering the 3D domain because of the artificial boundaries. The second issue has been also the subject of a review [31].
- In [31] we deal with numerical simulations of incompressible Navier-Stokes equations in truncated domain. In this context, the formulation of these equations has to be selected carefully in order to guarantee that their associated artificial boundary conditions are relevant for the considered problem. In this paper, we review some of the formulations proposed in the literature, and their associated boundary conditions. Some numerical results linked to each formulation are also presented. We compare different schemes, giving successful computations as well as problematic ones, in order to better understand the difference between these schemes and their behaviours dealing with systems involving Neumann boundary conditions. We also review two stabilization methods which aim at suppressing the instabilities linked to these natural boundary conditions.
- In [40], we propose a framework for Windkessel parameter estimation in a 0D representation of the 3D fluid-flow domain. Parameters are estimated from uncertain measurements through a sequential approach, and the 0D representation is iteratively improved through 3D-CFD simulations. The application of generalized sensitivity functions to assess parameter correlation and to ascertain the measurement set needed to avoid identifiability problems is also presented through representative test cases. This method, which is capable of handling non-simultaneous measurements, is demonstrated and validated for a patient-specific case of aortic coarctation.
- In [17] we perform the first patient-specific pulmonary hemodynamics 3D-0D modeling before single ventricle stage 2 surgery. 0D parameters are automatically tuned to match flow and pressure clinical measurements that are not taken where 3D boundary conditions need to be specified. This work on six patients demonstrates how simulations can help to check the coherence of clinical data or provide insights to clinicians that are otherwise difficult to measure, such as in the presence of kinks.

- In [25] we study a case of post single ventricle stage 2 surgery with the three following aims: (i) to show how to build a patient-specific model describing the hemodynamics in the presence of collaterals, using patient-specific clinical data collected at different times; (ii) to use this model to perform virtual collateral occlusion for quantitative hemodynamics prediction; and (iii) to compare predicted hemodynamics with post-operative measurements.

6.4. Numerical methods for cardiac electrophysiology

Participants: Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Elisa Schenone.

- In [33], a reduced-order method based on Approximated Lax Pairs (ALP) is applied to the integration of electrophysiology models. These are often high- dimensional parametric equation systems, challenging from a model reduction stand- point. The method is tested on two and three dimensional test-cases, of increasing complexity. The solutions are compared to the ones obtained by a finite element. The reduced-order simulation of pseudo-electrocardiograms based on ALP is proposed in the last part.
- In [21], we address the question of the discretization of Stochastic Partial Differential Equations (SPDE) for excitable media. Working with SPDE driven by colored noise, we consider a numerical scheme based on finite differences in time (Euler-Maruyama) and finite elements in space. Motivated by biological considerations, we study numerically the emergence of reentrant patterns in excitable systems such as the Barkley or Mitchell-Schaeffer models.

6.5. Lung and respiration modeling

Participants: Laurent Boudin, Muriel Boulakia, Céline Grandmont, Jessica Oakes, Ayman Moussa, Irène Vignon-Clementel.

- In [20], we consider the non-reactive fully elastic Boltzmann equation for mixtures. We deduce that, under the standard diffusive scaling, its limit for vanishing Mach and Knudsen numbers is the Maxwell-Stefan model for a multicomponent gaseous mixture.
- In [49], we first deal with the modelling and the discretization of an aerosol evolving in the air, in the respiration framework, within a domain which can be fixed or moving. We also investigate basic numerical properties of the numerical code which was developped, and also focus on the influence of the aerosol on the airflow.
- In [38], the aim of the study was to determine susceptibility differences between healthy and emphysematous rats exposed to airborne particles. To do this, we performed animal exposure experimenters and measured particle deposition concentrations with Magnetic Resonance Imaging. We showed that overall deposition was significantly higher in the elastase-treated rats compared to the healthy ones, suggesting enhanced susceptibility to airborne particles in diseased lungs. Current work aims at integrating such experimental data into modeling [39] and compare numerical simulations with experiments. To extend particle modeling to expiration, a 1D particle transport model is under development [44].

  While it is known that the retention of fine particles is less in microgravity (uG) compared to normal gravity (1G) levels, it was unknown the spatial relationship of deposited particles. In [26], rats were exposed to 1 micron diameter particles on the NASA uG airplane and compared to rats exposed in 1G. We found that the ratio of deposited particles in the central airways compared to the peripheral ones, was significantly less in the uG than in 1G, indicating enhanced deposition in the periphery. This data suggests that toxicology effects of exposure to Moon dust may not be insignificant.
- In [51], we establish stability estimates for the unique continuation property of the nonstationary Stokes problem. These estimates hold without prescribing boundary conditions and are of logarithmic type. They are obtained thanks to Carleman estimates for parabolic and elliptic equations. Then, these estimates are applied to an inverse problem where we want to identify a Robin coefficient defined on some part of the boundary from measurements available on another part of the boundary.

6.6. Miscellaneous

Participants: Jean-Frédéric Gerbeau, Damiano Lombardi, Marina Vidrascu.

- in [32] a reduced-order model algorithm, called ALP, is proposed to solve nonlinear evolution partial differential equations. It is based on approximations of generalized Lax pairs. Contrary to other reduced-order methods, like Proper Orthogonal Decomposition, the basis on which the solution is searched for evolves in time according to a dynamics specific to the problem. It is therefore well-suited to solving problems with progressive front or wave propagation. Another difference with other reduced-order methods is that it is not based on an off-line / on-line strategy. Numerical examples are shown for the linear advection, KdV and FKPP equations, in one and two dimensions.
- in [41] we propose a direct method for computing modal coupling coefficients - due to geometrically nonlinear effects - for thin shells vibrating at large amplitude and discretized by a finite element (FE) procedure. These coupling coefficients arise when considering a discrete expansion of the unknown displacement onto the eigenmodes of the linear operator. The evolution problem is thus projected onto the eigenmodes basis and expressed as an assembly of oscillators with quadratic and cubic nonlinearities. The nonlinear coupling coefficients are directly derived from the finite element formulation, with specificities pertaining to the shell elements considered, namely, here elements of the ”Mixed Interpolation of Tensorial Components” family (MITC). Therefore, the computation of coupling coefficients, combined with an adequate selection of the significant eigenmodes, allows the derivation of effective reduced-order models for computing - with a continuation procedure - the stable and unstable vibratory states of any vibrating shell, up to large amplitudes. The procedure is illustrated on a hyperbolic paraboloid panel. Bifurcation diagrams in free and forced vibrations are obtained. Comparisons with direct time simulations of the full FE model are given. Finally, the computed coefficients are used for a maximal reduction based on asymptotic nonlinear normal modes (NNMs), and we find that the most important part of the dynamics can be predicted with a single oscillator equation.
- in [53] we deal with the following data assimilation problem: construct an analytical approximation of a numerical constitutive law in three-dimensional nonlinear elasticity. More precisely we are concerned with a micro-macro model for rubber as the one proposed in [36]. Macroscopic quantities of interest such as the Piola-Kirchhoff stress tensor can be approximated for any value of the strain gradient by numerically solving a nonlinear PDE. This procedure is however computationally demanding. Hence, although conceptually satisfactory, this physically-based model is of no direct practical use. We aim to circumvent this difficulty by proposing a numerical strategy to reconstruct from in silico experiments an accurate analytical proxy for the micro-macro constitutive law.

SISYPHE Project-Team

5. New Results

5.1. Fault detection and localization in networks of transmission lines

Participants: Mohamed Oumri, Michel Sorine.

Some results have been obtained in collaboration with Florent Loete (LGEP) and Qinghua Zhang (Inria, I4S):
- *Experimental validation of the inverse scattering method for distributed characteristic impedance estimation.* Our theoretic results and numerical simulations have shown the ability of inverse scattering-based methods to diagnose soft faults in electric cables, in particular, faults implying smooth spatial variations of cable characteristic parameters. We have obtained laboratory experiments confirming the ability of the inverse scattering method for retrieving spatially distributed characteristic impedance from reflectometry measurements. Various smooth or stepped spatial variations of characteristic impedance profiles have been tested. The tested electric cables are CAN unshielded twisted pairs used in trucks and coaxial cables [37].

- *Diagnosis of networks using tagged electric lines.* A new electromagnetic marking method of transmission lines has been proposed for diagnosis of electric networks when conditions of uniqueness of the solution are not fulfilled (e.g. in case of symmetries): small non-interfering characteristic defaults are added to the lines and used as tags. A patent application has been submitted [36].

A new application of our monitoring technique has been explored in collaboration with EDF and a first result has been obtained:
- *Monitoring of post-tensioned ducts or water content in concrete walls with embedded transmission lines.* We have presented an electromagnetic method of diagnosis based on frequency domain reflectometry (FDR) associated with our inversion algorithm, ISTL (Inverse Scattering for Transmission Lines). ISTL allows one to estimate the spatial profile of the electrical impedance of the line from the FDR measurements. Experimental results on two mockups of external post-tensioned ducts with filling defects show the feasibility of the method. We will try to show the similarities between auscultation external post-tensioned ducts and measurement of water content by TDR probes (Time Domain Reflectometry) [34].

- *Fault diagnosis of wired electric networks by reflectometry.* A first extension to Baum-Liu-Tesche equations has been proposed in [31].

5.2. Cardiovascular signal processing and applications

Participants: Lisa Guigue, Claire Médigue, Michel Sorine, Serge Steer.

See the Sofware section 4.1 for a description of tools developed for *Cardiovascular Waves Analysis.*

5.3. Glycemic control in ICUs

Participant: Michel Sorine.

The results of statistical analysis of the data gathered during the large clinical trial CGAO-REA have been published in [14]: “Tight computerized versus conventional glucose control in the ICU: a randomized controlled trial”. Despite the increase in the incidence of severe hypoglycemia in our experimental group, based on the absence of difference in mortality between patients on tight computerized glucose control and those on less stringent glucose control without computerized decision support systems (CDSS), this study could pave the way for future randomized controlled trials assessing new generation CDSSs allowing the safe implementation of blood glucose control in the ICU that take into account the complexity of glucose control throughout the ICU stay and the variability of individualized insulin needs. Some new objectives for computer aided glycemic control in ICUs have been proposed in [32]. An article proposing a more detailed statistical analysis of the severe hypoglycemic events has been submitted.

5.4. Modeling and optimizing patient pathways in hospital

Participants: James Leifer, Michel Sorine.

External scientific collaboration with:
- Niccolo Curatolo, Directeur des opérations, Hôpitaux universitaires Paris-Sud, Assistance publique-Hôpitaux de Paris (AP-HP);
- Dr Maurice Raphaël, Chef de service, Urgences adultes, Hôpital Bicêtre, AP-HP;
- Dr Christophe Vincent-Cassy, Responsable des systèmes informatiques des urgences, AP-HP;
- Lucie Gaillardot-Roussel, Ingénieur en organisation, AP-HP;
- Dr Paul Jarvis, Senior consultant doctor in emergency medicine, Calderdale and Huddersfield National Health Service Foundation Trust, UK.

In 2014, we began a case study of the emergency department (ED) at Bicêtre Hospital, a large ED handling 50,000 patient visits per year, which is amongst the top 10 by volume and by annual volume growth for EDs in the Paris region.

Rather than presume the appropriateness of a predetermined scientific formalism, our strategy was to allow the application to frame a series of questions in order to lead us to experiment with several potential scientific tools at the present “low risk, high uncertainty” phase of investigation:

- *Top-down modeling:* Can we capture the expert knowledge of doctors and nurses as to the pathways followed by their patients by transforming this knowledge into a series of “use case” rules borrowed from the techniques of software specification? Can these rules by transformed into an executable model using business process modeling languages and tools (Orc, YAWL, ...) for simulating the complex parallel composition of man-machine processes in a hospital setting?

- *Bottom-up modeling:* How can the hospital be instrumented for cheaply and accurately capturing its real activity (movement of people and machines, delays, errors, ...) and tuning the parameters of the model? Can we intercept HL7 messages (a standardized electronic message format for medical data) and/or access raw time-stamped database entries to use machine learning techniques (particularly process mining) to extract from the running hospital the graphs representing the actual sequence of care events in order to get rapid feedback about the most heavily used and most often delayed path segments?

- *Underlying cost semantics:* Can we formalize in process calculi (for example, a variation of pi calculus) the “micro internal economy” of costs exchanged inside the hospital to quantify the economic performance of each patient pathway?

- *Offline experimentation and optimization:* Can potential optimization to the model be explored offline in a sort of “serious game” to allow non-intrusive experimentation with different strategies for eliminating bottlenecks, increasing flow rates, decreasing costs, etc.?

- *Data visualization for medical personnel:* Given that the medical personnel themselves are best suited to fixing the daily frictional time losses that most are resigned to accept as “part of the job”, how can the model be presented in a visually lucid manner to render the previously “invisible” aspects of the hospital’s organization visible?

- *Online real-time control:* Can the feedback loop be completed and the model be used to directly provide real-time visual feedback to the hospital personnel to enable them to measure their systemic progress (or systemic unintended consequences) of their localized optimizations?